FOOD STANDARDS AGENCY

RESEARCH PROJECT G03019

FINAL REPORT

January 2011

A PROTEIN-BASED INVESTIGATION OF THE USE OF MUDPIT AS A TOOL FOR THE SEPARATION AND QUANTIFICATION OF PROTEINS FOR GM CROP SAFETY ASSESSMENTS

Francesca P. Robertson, Paul D. Fraser, John M. Halket, Raj K.P. Patel, Kaisa Koistinen, Christopher Gerrish, Peter M. Bramley

School of Biological Sciences

Royal Holloway, University of London

CONTENTS

1 EXECUTIVE SUMMARY

Despite advances in gel-based approaches to proteomic analysis, the technology remains unsuitable for rigorous safety assessment of GM plants. Limiting factors include the type of proteins that can be analysed (i.e. hydrophobic proteins are poorly detected), inadequate dynamic range and inaccurate quantification. Multidimensional protein identification technology (MudPIT) is an attractive alternative. The procedure has been shown to detect and quantify low abundance peptides and peptides with hydrophobic properties.

The aim of the project was to assess the potential of MudPIT as a quantitative procedure to ascertain protein perturbations arising from genetic manipulation. The first stage of the project aimed to optimise front end procedures to maximise proteome coverage by MudPIT. The second stage of the project applied both so called‘non-chemical’ label-free methods, i.e., stable isotope labelled standards) and ‘chemical’ labelling methods (e.g. iTRAQ) to evaluate protein quantification of GM plant material. The unique collection of GM material and plants available at RHUL were used to fully evaluate and validate the procedure.

The MudPIT workflow was successfully developed for different mass spectrometry platforms: ESI Q-TOF, Orbitrap and Linear Ion traps and MALDI-TOF/TOF. A ‘non chemical’ label-free method was used to quantify iTRAQ labelled peptides using MS^3. MudPIT was used with iTRAQ to gain quantitative information for 150 protein perturbations in non-GM and GM tomato cultivars developed and grown at RHUL using the Agilent 6520, QTOF. In the azygous cultivar, 7 proteins showed significant difference

from the wild type; these were stress response proteins. In the GM *Psy1* sense cultivar, 60 proteins were found to be perturbed. 59 proteins were down-regulated and one protein was found to be significantly elevated in *Psy1* sense: abscisic acid stress ripening inhibitor protein 1. The gene product phytoene synthase from the intended genetic alteration in *Psy1* sense was notably absent from the iTRAQ quantitative protein profiles using the QTOF.

2 Glossary

LC MS	Liquid Chromatography Mass Spectrometry
LC ESI MS	Liquid Chromatography Electrospray Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
Q-TOF MS	Quadrupole Time Of Flight Mass Spectrometry
ESI	Electrospray ionization
FT	Fourier Transform
GM	Genetically modified
EPSPS	5-enolpyruvylshikimate-3- phosphate synthase
AZ	Azygous control
Psy1	Phytoene synthase
AC	*Ailsa Craig* parent background
MudPIT	Multi Dimensional Protein Identification Technology
iTRAQ	Isobaric Tags For Relative And Absolute Quantification
ICAT	Isotope Coded Affinity Tags
SCX	Strong Cation Exchange
SDS PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
2D-PAGE	Two Dimensional Polyacrylamide Gel Electrophoresis
CAN	Acetonitrile
FA	Formic Acid
BSA	Bovine serum albumin
TEMED	N,N,N’,N’-tetramethylethylenediamine
DTT	Dithiothreitol
HCl	Hydrochloric acid
CHAPS	3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate

3 Introduction

In Europe, the consumer has been reluctant to accept the introduction of GM/novel foods into the food chain. Perceived safety implications on human health are a contributing factor to this stance. Therefore, transparent, robust, unambiguous safety assessments are required to ease the general public's fears of GM foods and crops. A study conducted in the US on product attributes and consumer acceptance of nutritionally enhanced, genetically modified foods concluded that public trust and confidence in various private and public institutions are significantly related to their acceptance of food biotechnology (Hossain and Onyango, 2004). Consumer acceptance of genetically engineered foods is driven primarily by public perceptions of risks, benefits and safety of these food products.

The current use of substantial equivalence has failed to convince the consumer that GM foods are safe to eat (Herrick, 2005). The concept of substantial equivalence was developed as part of the safety evaluation framework and held the idea that existing foods were safe and could be used for comparing the properties of genetically modified foods. Substantial equivalence marks the starting point in the safety evaluation, rather than the end point. Application of substantial equivalence is not a safety assessment in its own right. Instead it helps to identify similarities and differences that exist between food in the market place and the new product. Once differences have been identified, they can be subjected to further toxicological investigation. Development and validation of new profiling methods for the identification and characterization of unintended effects, which may occur as a result of genetic modification, are recommended

(Kuiper *et al.*, 2001).

The success of the concept of substantial equivalence depends on three critical elements:

- The availability of an appropriate comparator and an understanding of the range of variation to be expected within the measured characteristics of that comparator.
- The choice of parameters in the single constituent compound analyses, the number and type of which will strongly influence the validity of any conclusions on comparative safety.
- The ability to discriminate between differences in the GM crop and the comparator that result from genetic modification and those differences in the plant's germplasm, some of which may be attributed to somaclonal variation introduced during tissue culture, and environmental or cultivation conditions (König *et al.*, 2004).

Critics of substantial equivalence claim that current testing approaches do not sufficiently address putative unintended and unexpected effects and cannot rule out the occurrence of potentially long-term effects that result from sustained human exposure to crops which might have subtle compositional changes that are difficult to detect (Millstone *et al.*, 1999). Despite criticism, the Food and Agriculture Organization of the United Nations (FAO) and World Health Organisation (WHO) conclude that it is the best available assessment paradigm and the concept for substantial equivalence is widely accepted internationally as the best guidance for safety assessment of new GM crops.

Through FSA and EU programmes, the suitability of new analytical

"omic" technologies such as metabolomics, transcriptomics and proteomics have been evaluated as techniques involved with the assessment of equivalence. A metabolomic study of field grown wheat concluded that the environment affects the metabolome and that any differences between the control and transgenic lines were within the same range as the differences observed between the control lines grown on different sites and in different years (Baker *et al.,* 2006). Another metabolomics study, using hierarchical metabolomics, demonstrated substantial compositional similarity between genetically modified and conventional potato crops (Catchpole *et al.,* 2005). A recent transcriptomics study on substantial equivalents in maize by Coll *et al.,* *(*2010) concludes that natural variation explains most transcriptomic changes between GM maize and their comparable non-GM varieties, when subjected to N-fertilisation farming practices.

The potential of these technologies to improve significantly our understanding of the metabolomic perturbations that occur in GM plants has been demonstrated, enabling more accurate and comprehensive comparisons with the non-GM equivalent lines to be made. Incorporation of these procedures into the safety assessments will provide rigour and add relevant data to assist in reassuring the public and accurately establish the changes that occur during genetic modification.

Significant advances have been made in the fields of metabolomics (Halket *et al.,* 2005) and transcriptomics (e.g. Affymetrix™ chips). The robustness and reliability of these technologies have reached an acceptable level for incorporation into safe assessment protocols for GM foods (Cellini *et al.,* 2004). Workflows for statistical analysis have been developed concurrently to cope with the massive data sets generated by such

experimentation. However, despite advances in proteomics, limitations remain in gel-based approaches. This puts into question their effectiveness as accurate methodologies for determining changes of proteins in GM material.

Traditionally, 2-D polyacrylamide gel electrophoresis (2D-PAGE) has been used to resolve complex mixtures of proteins prior to identification and quantification. Improvements have been made, such as using immobilised pH gradients, which allow hydrophobic (membrane) and basic proteins to be resolved, thus improve reproducibility and practical handling. 2D-PAGE has a poor dynamic range, preventing the detection of low abundance proteins, a limiting load capacity and artefacts due to chemical modification of the proteins. The procedure is not amenable to high throughput or automated analysis. It is time consuming, due in part to the necessity to run many replicate samples simultaneously to obtain reproducible data. Gel-based proteomics approaches, therefore, are not presently suitable for the reliable analysis of potential perturbations arising in GM plants.

Multidimensional protein identification (MudPIT) represents an attractive alternative to gel based proteomics and has been used for rapid, large scale proteome analysis (Wolters *et al.*, 2001). The technology has been shown to be unbiased and allows identification of low abundance hydrophobic proteins. In one study alone, over 1,484 proteins were detected within +/- 0.5 % between duplicate analysis and a dynamic range of 10,000 to 1 between the most abundant and the least abundant proteins/peptides in a complex mixture. Comprehensive exploration of the rice proteome revealed that MudPIT provided a larger coverage of proteins than 2D-PAGE, with

2363 proteins identified by MudPIt and 556 proteins identified by 2D-PAGE (Koller *et al.*, 2002). MudPIT can identify more membrane proteins and proteins of low abundance than 2D-PAGE.

The MudPIT technique coupled with fast scanning, a high mass accuracy MS platform has grown in stature amongst the scientific community. The technique has been applied to animal, plant and yeast systems extensively. In 2006 Chen *et al.*, were able to characterise over 1,000 proteins from large scale protein expression analysis of a metastatic breast cancer cell line, BCM2. The 1,000 proteins were identified from 11 collected fractions by MudPIT.

MudPIT experiments have been used to target proteins from subcellular fractions. By using sub fractionation prior to analysis the proteome coverage can be expanded, identifying proteins at low abundance. Cellular sub fractionation followed by MudPIT analysis has led to identification of proteins in exosomes from saliva in a study undertaken in 2009 by Gonzalez-Begne *et al*. This study identified 491 proteins from the exosome; the discovery of this information may be useful in the diagnosis and treatment of systemic diseases such as diabetes, prion disease and cancers

The principal contributing factors that have facilitated the emergence of protein/peptide analysis have been developments in modern mass spectrometry (MS) and the accessibility to DNA sequences. It is now possible to detect and identify peptides from protein digests at 10^{-8} mol of a protein.

Quantification is the primary challenge facing proteomics today.

MudPIT condenses the global quantification of complex protein mixtures into one experiment. Quantification strategies have been developed and implemented to determine the quantity of proteins present in a sample (Pan *et al.*, 2009). Absolute and relative quantification strategies have been developed. Absolute quantification methods use synthetic peptides to mimic endogenous target peptides in the sample. These peptides have the same physico-chemical properties, but contain a stable isotope labelled with ^{13}C or ^{15}N in an amino acid.

Absolute quantification of peptides can be achieved by using the Absolute Quantification (AQUA) method. AQUA uses stable-isotope labelled peptides synthesised from copies of corresponding target peptides with incorporated stable isotopes. These synthetic peptides are chemically identical to the target peptide. AQUA peptides are used to precisely quantify the absolute levels of each peptide using selected-reaction monitoring analysis with tandem spectrometry (Gerber *et al.*, 2003). Ocana *et al.*, 2007 used AQUA peptides to quantify EPSPS peptides in GM soya.

Strategies such as multiple reaction monitoring (MRM) on peptides can be used in conjunction with triple quadrupole machines and corresponding technology to get a relative quantification of peptides. MRM improves sensitivity and detection of peptides, which is very important in analysing peptides with modifications such as hydroxyproline and hydroxylysine residues (Wolf-Yadlin *et al.*, 2007).

Classical non-chemistry approaches have compared peak areas between chromatograms and used comparisons with peak areas derived from spiked synthetic peptides creating relative levels. Of the chemistry methods, Isotope Coded Affinity Tags (ICAT reagent is the best characterised. This

procedure works on the mass difference concept between two samples (2-plex) (Gygi *et al.*, 1999).

Although this approach has found useful applications, a number of disadvantages are evident, such as MS complexity, limitation to cysteine containing proteins and it is not amenable to identification of post translational protein modifications. ICAT has now been surpassed by iTRAQ (isobaric tags for relative and absolute quantification), which represents significant advancement in protein quantification by MS (Ross *et al.*, 2004). This chemical quantification method has been developed for peptides by Applied Biosystems. The iTRAQ chemical tags and theory are detailed in Figure 3.1 iTRAQ uses 4 identical chemical tags, with the same overall mass, to multiplex 4 samples for relative peptide quantification (Boehm *et al.*, 2007). Each label has an isobaric tag of 145Da, which consists of a balancer group and a reporter group. A fragmentation site is formed between the balancer and the reporter group. The label also has a peptide reactive group (NHS-ester) which forms specific attachments to free primary amine groups on N-termini of peptides and ε-amino lysine side chains. Peptides are labelled at the N-terminus following protein digestion (Fenyö and Beavis, 2008). When iTRAQ labelled peptides are fragmented using mass spectrometry, peaks from singly charged reporter group fragments appear in the m/z range of 114 to 117. Peptides are quantified by interpretation of the ratios of these fragment peaks (Boehm *et al.*, 2007).

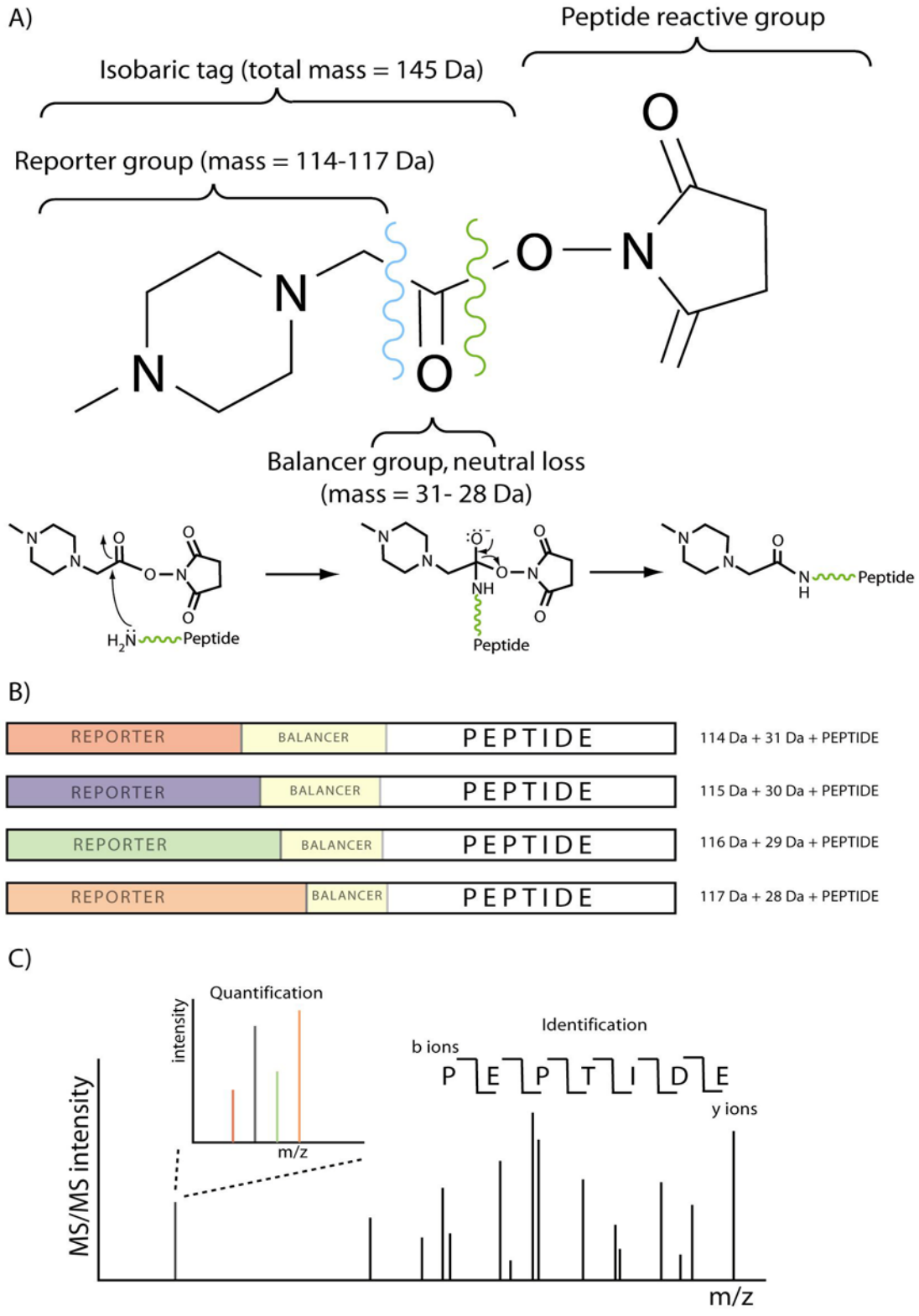


Figure 3.1 Isobaric tags for relative and absolute quantification of peptides (iTRAQ) method A) shows the structure of the iTRAQ tags and the reaction mechanisms. B) Components of the 4 isobaric tags displaying the reporter, balancer tags and the peptide.C) MS/MS can be used to identify peptides from b and y ions and ratios of up to 4 different samples tagged with iTRAQ tags can be compared to quantify peptides (Boehm et al., 2007; Moffitt Cancer Center and Research, 2009).

The aim of this project is to deliver a multiplexed quantitative MudPIT procedure to assess substantial equivalence in GM tomato cultivars. Such an outcome will represent a significant advancement to the state of the art. To reach this goal the front end of the MudPIT technology will be optimised. Previous work with membrane proteins and low level GM proteins has shown that initial enrichment or removal of abundant proteins can dramatically improve the diversity of proteins detected.

A key advancement will be the use of iTRAQ. This system represents a novel labelling strategy that can overcome some of the limitations associated with protein quantification. Lower cost alternatives, such as using stable isotopes with ion trap MS will be evaluated. Currently iTRAQ is the market leader and the multiplexing feature makes it ideal for the relative quantification of complex proteins mixtures, and suitability for the assessment of GM crops. Through industrial collaborations we have this technology functioning at RHUL. Finally the unique collection of GM tomato plants and soya material will provide audited samples for evaluation of the method.

4 Experimental procedures

4.1 General reagents and standards for proteomic investigations

Tris(hydroxymethyl)methylamine (Tris-HCl), analytical grade methanol, acetone, ethanol, HPLC-grade acetonitrile (ACN), hydrochloric acid (HCl), trifluoroacetic acid (TFA) and formic acid (FA) for protein sequencing analysis were purchased from VWR (Poole, UK). Bromophenol blue, sodium dodecyl sulphate (SDS), iodoacetamide, dithiothreitol (DTT), N,N,N',N'-tetramethylethylendiamine (TEMED), ammonium persulphate, ammonium bicarbonate, urea, thiourea, glycerol, bovine serum albumin (BSA) and Proteosilver™ plus silver stain kit were from Sigma (Poole, UK) Modified trypsin (sequencing grade) was purchased from Roche Diagnostics (Lewes, UK). Peptide standards for Q-TOF MS calibration were purchased from Fluka BioChemica (Poole, UK). Ultrapure protogel acrylamide and concentrated 10 X Tris/glycine/SDS (electrophoresis grade) were purchased from National Diagnostics (Hessel, UK). Protein standards for SDS-PAGE were purchased from Sigma (Poole, UK). Strong Cation exchange cartridges and iTRAQ kits for quantitative proteomic investigation were purchased from ABI, (Warrington, UK). Bradford protein assay reagent was from Bio-Rad (Hemel Hempstead, UK). C_{18} Zip Tip pipette tips were from Millipore (Watford, UK). Syringe filters and nitrocellulose filters (0.2 μm) were obtained from Whatman (Brentford, UK). Solid phase extraction C_{18} cartridges were purchased from Waters Ltd (Elstree, Hertfordshire). Synthetic peptide standards were synthesised by Peptide Protein Research

Ltd (Fareham, UK).

4.2 Generation, supply and preparation of genetically modified GM food crop.

A supply of GM and non-GM soya beans was obtained for the use in development of the quantitative MudPIT system. Ripe tomato fruit from GM lines and their background controls were collected 7 days post breaker from stable GM tomato varieties phytoene synthase 1 (*Psy1*) antisense, *Psy1*Sense over expressor, azygous controls and *Ailsa Craig* parent background. 3 fruit were taken from 3 plants from each line. Fruit were lyophilised and homogenised as in 4.3.2.

4.3 Protein extraction

4.3.1 Total protein content by Bradford assay

Protein concentrations of tomato and soya extracts were estimated using a Bradford-based assay from Bio-Rad. Standard protein solutions of bovine serum albumin (BSA) were prepared in deionised water at concentrations of 0.5, 2, 5, 10, 15, 20, 25 mg/ml. Generally, 100 µl of BSA solution was mixed with 750 µl of water and 150 µl of Bio-Rad reagent and incubated at room temperature for 3 min. Absorbance measurements from protein samples were taken at 595 nm and compared against the standard BSA calibration curve.

4.3.2 Preparation and extraction of proteins from lyophilised tomato and soya powder

Tomato and soya powder were weighed into 20 mg aliquots in

triplicate. Extraction buffer (500 µl; 7M urea, 2M thiourea, 2% CHAPS, 50 mM Tris- HCl, pH 8) was added and left at room temperature for 30 min. Tubes were centrifuged at 14,000 g for 10 min and the supernatant collected. Ice cold acetone (2.5 ml) was added and the mixture left at -20^{o}C overnight. Tubes were centrifuged at 14,000 g for 10 min and the supernatant discarded. Pellets were washed three times with acetone and dried in a vacuum centrifuge. Pellets were resuspended in 10 mM TEAB buffer and protein concentrations were obtained by Bradford assay (4.3.1). 100 µg protein aliquots were collected and dried down under vacuum centrifuge and stored at -80^{o}C prior to iTRAQ analysis. Details for homogenisation, lipid removal and buffer choice are in the Appendix.

4.4 SDS PAGE

Mini SDS PAGE gels (8 x 10 cm) were prepared using a Bio-Rad (Hemel Hempstead, UK) system employing 1.5 mm spacers and 10 well combs. Aliquots of 10 µg protein were loaded onto SDS-PAGE (resolving gels of 10, 12 or 15 % w/v) with a constant current of 15 mA. Gels were washed and silver stained according to the manufacturer's recommendation (Sigma, Proteosilver™ stain kit) or with Coomassie blue. Proteins were fixed in the gel using 100 ml solution of ethanol, acetic acid and water, (50:10:40 v/v/v) for a minimum of 40 min. Fixing solution was decanted and gels washed with 100 ml ethanol: water, (30:70 v/v) for 10 min and placed into 100 ml sensitizer solution (10 min). Gels were washed twice with 200 ml deionised water for 10 min before addition of silver staining solution (100 ml) for 10 min. Gels were then washed with deionised water for 1.5 min and placed in developing solution. The required intensity of silver

staining was reached in 5 - 7 min, and the reaction halted with addition of the Sigma stop solution.

4.5 OFFGEL and Rotofor- Protein separation

2.5 mg of tomato and soya protein extract were fractionated by isoelectric focusing using an Agilent 3100 OFFGEL fractionator following the manufacturer's instructions. Proteins were focused based on their pI using 24 cm long Immobiline DryStrip (GE Healthcare) with a linear pH gradient ranging from 3 - 12. 140 µl of protein extract was loaded in each well. After focusing, fractions were collected and stored at - 20 °C for further use. Target OFFGEL protein fractions were also directly subjected to trypsin digestion. A total of 20 µl of each OFFGEL fraction was mixed with 20 µl of 0.1% TFA. After homogenisation, samples were desalted using Zip-tips. Protein elution was performed with 20 µl of 70% ACN/30% 0.1 TFA. The solvent was evaporated using a GenVac evaporator and reconstituted with 15 µl of 50 mM ammonium bicarbonate, pH 8.0, followed by the addition of 15 µl of 12.5 ng/µl trypsin dissolved in the same buffer. Samples were incubated at 37ºC overnight, and then evaporated using the rotary evaporator. Reconstitution was carried out with an appropriate volume of 0.1% TFA in order to analyse the tryptic peptides. Details of use for the Rotorfor are found in the Appendix.

4.6 Proteolytic in-solution digestion

Protein concentration of tomato and soya were determined by Bradford protein assay. Soya and tomato proteins (100 µg) were reduced and

alkylated using the iTRAQ kit as to the manufacturer's instructions. In 100 µl, 10 mM TEAB solutions were digested with trypsin to create a protein to trypsin ratio of 1/100 to 1/1000 v/v. The digests were incubated at 37 °C overnight. Following digestion, tubes were placed at -20 °C before labelling with iTRAQ reagent.

4.7 iTRAQ labelling

Aliquots of digested tomato and soya (100 µg protein) were labelled with 4 multiplex iTRAQ reagents to the manufacturer's instructions. Proteins were reduced and alkylated before digestion with trypsin, Glu-C or Lys-N. Following labelling, aliquots were pooled for strong cation exchange (SCX) separation.

4.8 Strong cation exchange fractionation

Tomato and soya iTRAQ labelled peptides were pooled and fractionated using strong cation exchange (SCX) prior to nano C_{18} reverse phase LC. Peptide separations were performed using strong cation exchange cartridges (Applied Biosystems, Warrington). 100 µl of loading buffer (10 mM K_2HPO_4:ACN 75:25 v/v, pH 3) was added to trypsin digests and balanced to pH 3 by H_3PO_3. Strong cation exchange cartridges were conditioned with loading buffer. Peptide digests at pH 3 were loaded onto the cartridge. A stepped mobile phase gradient (50 mM, 100 mM, 150 mM, 200 mM, 250 mM, 500 mM K_2HPO_4:ACN 75:25 v/v, pH 3) was passed through the column and eluent collected from each fraction. Fractions were dried in a Genevac EZ-2 Evaporator vacuum centrifuge and stored at -20 °C.

4.9 LC-MS

4.9.1 Nano-LC ESI Q-TOF MS (Q-STAR, ABI)

Nano-LC ESI MS/MS experiments were performed on a QSTAR Pulsar I (Applied Biosystems, Warrington, UK) hybrid quadrupole time of flight mass spectrometer connected to a nano-LC system (LC Packings, Camberley, UK). Samples were loaded onto a 200 µm i.d. x 5 mm PS-DVB monolithic (Dionex, Sunnyvale, CA, USA) trap column with a flow rate of 10 µl/min of 0.1 % TFA for 30 min. After pre-concentration, the trap column was automatically switched in-line with the PS-DVB monolithic (3 µm, 100 µm id. x 50 mm, Dionex) analytical column and the peptides eluted with a linear gradient starting at 95 % eluent A (0.1 % v/v formic acid in water) to 40 % of eluent B (0.1 % v/v formic acid in ACN) at 40 min, the flow rate being 300 nl/min. HPLC fractions of tryptic peptides were injected using an LC Packings FAMOS autosampler and UltiMate LC pumps. A Protana nanospray interface and 10 µm distal coated fused silica PicoTips (New Objective, Woburn, USA) were used for the nanoESI.

The positive TOF mass spectra were recorded using information-dependent acquisition (IDA). TOF MS survey scans were recorded for mass range m/z 400 to 1600, followed by MS/MS scans of the two most intense peaks. Typical ion spray voltage was in the range of 2.0 to 2.4 kV and N_2 was used as the collision gas. Other source parameters and spray positions were optimised with the tryptic digest of bovine serum albumin. Analyst QS 1.0 sp8 software from Applied Biosystems was employed for data analysis.

4.9.2 nanoLC ESI Q-TOF MS (Agilent 6520)

Nano-LC ESI MS/MS experiments were also performed on an Agilent QTof 6520 mass spectrometer hybrid quadrupole time of flight mass spectrometer connected to a Agilent 1200 nLC (Agilent, Wokingham UK) nano-LC system. 50, 100 & 150 mM cation exchange fractions were resuspended with 50 µl 1% v/v formic acid. Due to the amount of salt present, the 500 mM samples were resuspended with 75 µl 1% v/v formic acid. A 10 µl injection was loaded onto a Agilent Chip Cube trap column with a flow rate of 4 µl/min of 0.1 % v/v TFA for 5 min. Chromatographic separations were made using the Agilent Chip Cube (Large Capacity Chip (II) G4240-62010, Separation: 150 mm x 75 µM, Enrichment: 9 mm 160 nl, Zorbax 300SB-C18 5 µM). An isocratic gradient, with a flow rate of 0.3 µl/min of solvent A (0.1 % FA in H_2O) for 3 min followed by a linear gradient from 5 - 45 % solvent B (0.1 % FA in 90 % ACN) for 128 min was used. Positive TOF mass spectra were recorded using information-dependent acquisition (IDA). TOF MS survey scans were recorded for mass range m/z 400 to 1600 every 200 ms, followed by MS/MS scans of the four most intense peaks every 250 ms. RAW files were extracted with Protein Hunter software (Agilent) into Mascot generic files (.mgf) prior to merging and analysis by Mascot 2.3 (Matrix Science) to generate iTRAQ and identification data

4.9.3 nanoLC ESI Orbitrap FTMS (LTQ Orbitrap XL, Thermo)

LC-MS analyses were carried out using a Surveyor LC system (Thermo Electron Corp, San Jose, CA USA.) coupled directly to a LTQ Orbitrap XL (Thermo Electron Corp, San Jose, CA USA). Samples were

loaded onto a Michrom CapTrap trap column, with a flow rate of 5 µl/min of 0.1 % TFA for 5 min. After pre-concentration, the trap column was automatically switched in-line with the Michrom C_{18} column 100 x 0.1 mm i.d analytical column and the peptides eluted with a linear gradient from 5 - 40 % solvent B (0.1 % v/v FA in ACN) for 100 min. Positive mass spectra were recorded using data-dependent MS/MS. Full MS survey scans were recorded for mass range m/z 400 to 1600 at 30,000 resolution, followed by MS/MS scans of the three most intense peaks using HCD at 7,500 resolution.

4.9.4 Capillary LC- ESI Ion Trap (LCQ Deca)

LC-MS analyses were carried out using an AS3000 autosampler, a P4000 gradient LC system (Thermo Electron Corp, Jose, CA USA..) coupled directly to a LCQ Deca ion trap (Thermo Electron Corp, San Jose, CA USA). 30 µl of each sample were injected into the LCMS system by using the autosampler.

Chromatographic separations were made using CLIPEUS C_{18} column 150 x 0.5 mm i.d. (Higgins Analytical Inc, Mountain View CA,USA.), using the following conditions. An isocratic gradient with a flow rate of 18 µl/min of solvent A (0.1 % v/v FA in H_2O) for 30 min followed by a linear gradient from 1 - 40 % solvent B (0.1 % v/v FA in ACN) for 120 min. Operating conditions for the ion trap mass spectrometer were: ESI positive mode, capillary temperature 220 ºC, isolation with 3.0Da collision energy normalized to 33 %, nitrogen sheath gas flow 34, no auxiliary gas, source voltage 4.5 kV and capillary voltage 3.0 V. The first scan event was full MS scan from 300 - 2000 m/z. The total number of microscans was 3 in 400 ms.

The second scan event was dependent MS/MS scan of the most intense ion enabling dynamic exclusion (after 3 scans of the most intense ion). The total number of MS/MS microscans was 3 in 600 ms.

Data acquisition was performed using Excalibur v1.2 software. Peptide identification was done by interrogating the generated .dta files containing MS/MS data into MASCOT search engine (http://www.matrixscience.com), against the NCBInr database using a mass tolerance of 0.6 Da.

4.9.5 MALDI-TOF-TOF MS/MS with nanoLC spotter

4.10 Cation exchange fractions were resuspended in 100 - 200 µl 0.05 % v/v TFA and 2 µl injected. Nano-LC analyses were carried out using an EASY-nanoLC (Bruker Daltronics, Coventry, UK). Chromatographic separations were made using a Nanoseparations C_{18}, 20 x 0.1 mm trapping column, followed by a PepMap C_{18}, 150 x 0.075 mm (Dionex, Sunnyvale, CA, USA). An isocratic gradient, with a flow rate of 0.3 µl/min of solvent A (0.05 % v/v TFA in H_2O) for 5 min, followed by a linear gradient from 2 - 45 % solvent B (0.05 % TFA in 90 % ACN v/v) for 64 minutes. A Proteineer FC II fraction collector (Bruker Daltronics, Coventry, UK) was used to spot 384 fractions, 10s each on a MALDI target plate MTP AnchorChip 800µm-384 (Bruker Daltronics), using HCCA as the matrix. An Autoflex TOF/TOF MS (Bruker Daltronics, Coventry, UK) was used in positive mode to acquire MS spectra from 1000 laser shots and MS/MS

spectra were acquired from 1000 laser shots per MS/MS spectrum. Database searches were performed with MASCOT 2.2 (Matrix Science) using the NCBInr database for viridiplantae (green plants) with trypsin (1 missed cleavage), Methylthio (C), Ox (M), iTRAQ4plex (K), iTRAQ4plex (N-term), with a MS tolerance of 30 ppm and MS/MS tolerance of 0.5 Da. The database search results of all four cation exchange fractions were compiled by ProteinScape (Bruker Daltronics, Coventry, UK). Protein search engines through Bruker Daltronics software were used on Swiss Prot, NCBI nr and MSDB (.http://www.matrixscience.com/help/database_help.html). Statistical methods and multivariate data analysis

Multivariate methods can be used to investigate the relations between all variables in experiment in a single context, allowing data to be displayed in scatter plots. Methods for processing multivariate data include principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) modelling. Multivariate data analysis was performed with SIMCA software (version 12, Umetrics, Umeå, Sweden). PCA, an unsupervised method, was carried out, primarily, to detect outliers. The supervised method, PLS-DA, was performed on data sets to identify which variables were responsible for sample classification. The students *t*-test and correlation analysis were performed using Excel.

4.10.1 Students *t*-test

The students t—test was calculated in Excel software, Microsoft.

4.10.2 PCA

PCA takes a data matrix with X and Y variables and calculates vectors of scores (t) by taking into consideration all variables entering the analysis. Two PC score vectors (t1 and t2) were plotted against each other; examples of PCA score plots are displayed in Section 5, Figure 5.21. PCA loadings plots show the relationship between X variables and Y variables. PCA provides loading vectors (p), showing how variables are combined to form the scores. Loadings correlation score vectors (p(corr)1 and p(corr) 2) can be used to see which variables are important and correspond to the directions in the score plots as shown in Figure 5.21 (B).

4.10.3 PLS-DA

PLS-DA models can be used to express a set of X-score vectors, Y-score vectors, X-weight vectors and Y-weight vectors. A linear relationship, dimension ‘a’, is expressed between X-score vectors ‘ta’ and Y-score vectors ‘ua’. The weight vectors of each model express how X-variables are combined to form ta-values and how the Y-variables are combined to form ua-values. PLS analysis generates weight coefficients from the variables; weights for the X-variables ‘w’ indicate the importance of these variables and how they associate with the modelling of Y. The weights for Y-variables ‘c’ show which Y-variables are modelled in respective PLS model dimensions. The ‘wc’ coefficient plot, shows the relationships between X and Y variables. PLS-DA plots are shown in Figure 5.22.

Loading plots were used for selection of potential biomarkers, which were confirmed by t-tests. PLS-DA can be used for class prediction of samples not included in the original model. PLS methods were applied for the pairwise analysis of different PAP species, for biomarker mining and

class membership prediction of new samples.

5 Results - Optimisation of front end procedures

5.1 Generation, supply and preparation of genetically modified GM plants

A supply of GM and non-GM soya beans was obtained for the use in development of the quantitative MudPIT system. GM and non-GM tomatoes were planted and grown at RHUL. Table 5.1 and Figure 5.1 display the tomato cultivars chosen, any genetic modification and the phenotypic effect.

Table 5.1 Tomato cultivars used to assess protein perturbations using MudPIT

Cultivar	GM modification	Phenotype
Ailsa craig	None	Wild type
Azygous	Been through the transformation process with no transgene incorporated	Identical to *Ailsa craig*
CRT I	Bacterial phytoene desaturase (Crt I)	Elevated levels of β-carotene
Psy1 Antisense	Down regulation of *psy1* in fruit	Lower levels of carotenoids
Psy1 Sense	Up regulation of *psy1* in fruit	Higher levels of carotenoids

Ripe fruit from all lines were harvested 7 days post breaker. Three fruit from each line were pooled from three plants of each cultivar. Seeds were removed from the fruit before pooling. Fruits were chopped and frozen in liquid nitrogen prior to freeze drying. Tissues was homogenised with a tissue lyser to powder for protein extraction. Figure 5.2 displays the workflow for collection and preparation of tissue for protein analysis.

AC-Wild type

Azygous

Down-regulated *Psy-1*
NO carotenoids

Crt I cultivar - high pro-vitamin A varieties

Figure 5.1 Tomatoes varieties generated and cultivated at RHUL

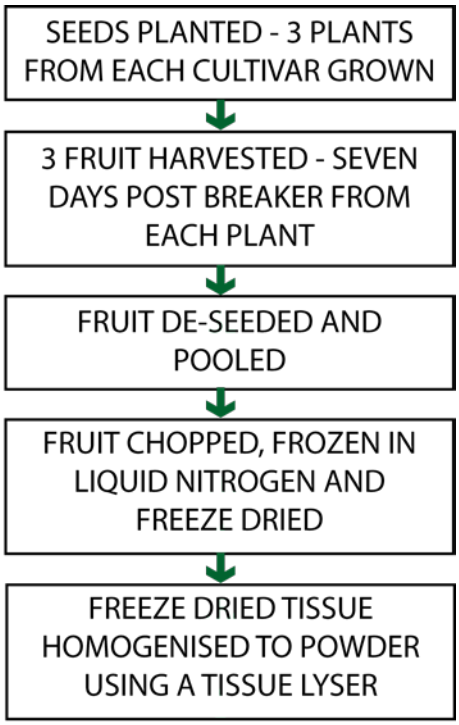


Figure 5.2 Workflow for collection and preparation of tomato fruit for protein analysis

5.2 Optimisation of protein extraction from soya and tomato

Plant material homogenised by freezer mill gave the most effective homogenisation method. Different solutions and extraction conditions were optimised to maximise protein solubilisation. In order to extract basic and membrane proteins, buffers with detergents and denaturing agents were tested. Good solubilisation was achieved using a solution of 7M urea, 2 M thiourea, 2% CHAPS and 10 mM DTT for 30 min with shaking at room temperature. Figure 5.3 displays non-GM soya extracts after 15, 30 and 60 min with urea based buffer.

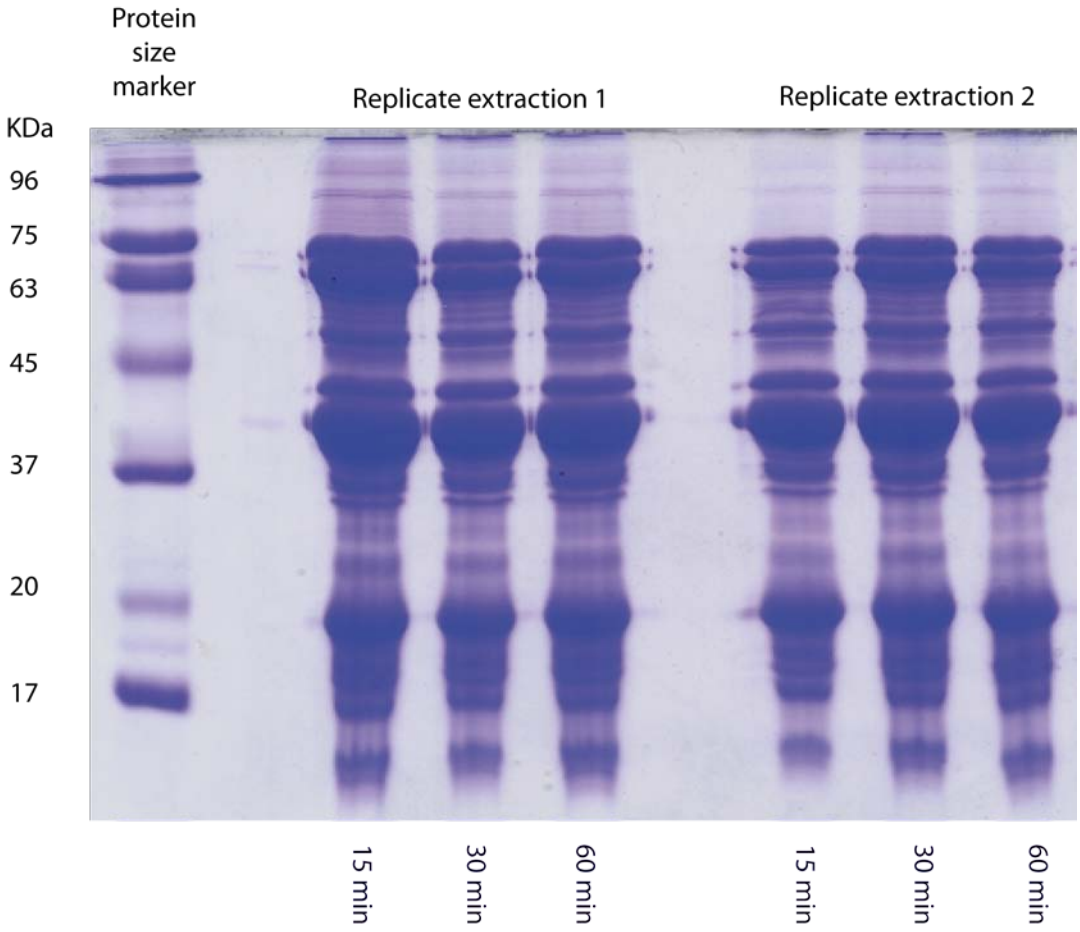


Figure 5.3 SDS PAGE of non GM soya extracted with 7M urea, 2M thiourea, 2 % CHAPS, stained with Coomassie blue.

5.3 Development of purification procedures of the protein extract as a means of improving proteome coverage by MudPIT

Membrane and soluble fractions were separated using ultra centrifugation at 105,000 g. Liquid isoelectric focusing (Rotorfor, BioRad) enriched proteins and decreased the complexity of the tomato samples. However, the most abundant proteins were found in several fractions and dominated the enrichment (Figure 5.4). This method would not be suitable for high throughput of analysis, as the subsequent MS analysis is too costly.

The Agilent OFFGEL system was trialled with tomato samples. This method is costly and increases the number of samples for analysis. It would not be suitable for high throughput of analysis.

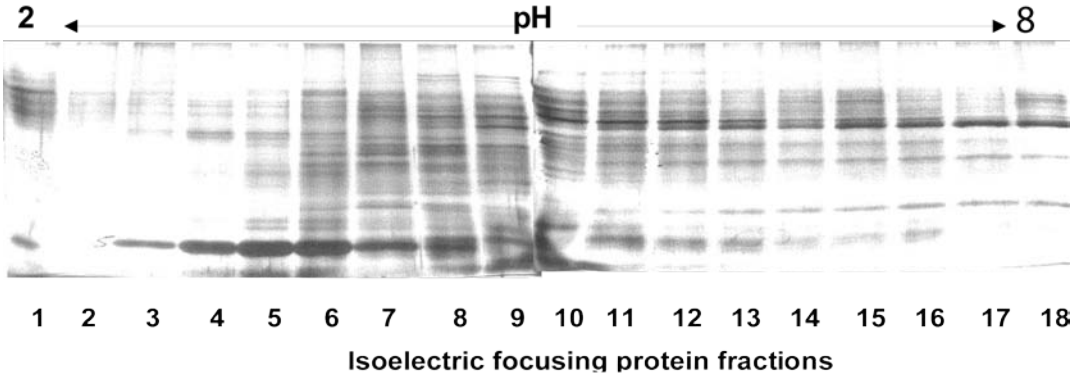


Figure 5.4 Isoelectric focusing of soluble tomato protein extraction using Rotofor.

5.4 Development of proteolytic digestion procedures for complex protein mixtures that are compatible with quantitative MudPIT approaches

Three proteolytic enzymes (trypsin, Glu-C and Asp-N) were tested to obtain the optimal digestion of tomato and soya proteins for MudPIT.

Trypsin was found to be the most effective digestion enzyme, because it gave the greatest sequence coverage and identified 97 proteins in tomato. Table 5.2 presents data of the total number of identified proteins and protein coverage of selected proteins digested with the different proteases. The results indicate that different enzymes give complementary results.

Table 5.2 Protein in-solution digestions with trypsin, Glu-C and Asp-N solubilisation to 0.1 M TEAB, pH 8.5 1 µg enzyme to digest 75 µg of protein overnight at +37°C

Protein	Database accession number	Coverage (trypsin)	Coverage (Glu-C)	Coverage (Asp-N)
Tomato		**97 proteins**	**11 proteins**	**13 proteins**
abscisic stress ripening protein 1	gi|584786	(18/115 = 15.7%)	(33/115 = 28.7%)	(0/115 = 0%)
coat protein	gi|229181	(120/158 = 75.9%)	(42/158 = 26.6%)	(32/158 = 20.3%)
acid beta-fructofuranosidase precursor (Acid suc rose hydrolase)	gi|124701	(131/636 = 20.6%)	(29/636 = 4.6%)	(163/636 = 25.6%)
1-aminocyclopropane-1-carboxylate oxidase homolog (Protein E8)	gi|119640	(94/363 = 25.9%)	(0/363 = 0%)	(115/363 = 31.7%)
Soya		**32 proteins**	**14 proteins**	**13 proteins**
glycinin G4 subunit	gi|255224	(187/560 = 33.4%)	(0/560 = 0%)	(0/560 = 0%)
alpha subunit of beta conglycinin [Glycine max]	gi|9967361	(126/559 = 22.5%)	(170/559 = 30.4%)	(99/559 = 17.7%)
sucrose-binding protein precursor (SBP)	gi|548900	(54/524 = 10.3%)	(66/524 = 12.6%)	(0/524 = 0%)
24 kDa oleosin isoform [Glycine max]	gi|266689	(13/223 = 5.8%)	(28/223 = 12.6%)	(10/223 = 0%)
P34 probable thiol protease precursor	gi|129353	(0/379 = 0%)	(33/379 = 8.7%)	(19/379 = 5.0%)

5.5 Optimisation of first and second dimension chromatographic separations- Establishment of high resolution, reproducible off and on line multidimensional separations for both ESI-qTOF MS/MS and ESI-MS^n (ion trap)

An offline cation exchange method was developed and optimised for both tomato and soya proteins. Good separation of peptides was obtained (Figure 5.5). A strong cation exchange cartridge system was trialled and separated peptides prior to RP nanoLC. Offline peptide separation using Agilent Offgel system enabled identification of membrane and basic peptides. However, abundant proteins were found across multiple fractions and dominated the profile. Table 5.3 details the number of proteins characterized from cation exchange HPLC, SCX cartridge system and

Agilent Offgel fractionation using two different databases, ie Viridiplantae database from NCBI and NRDB.

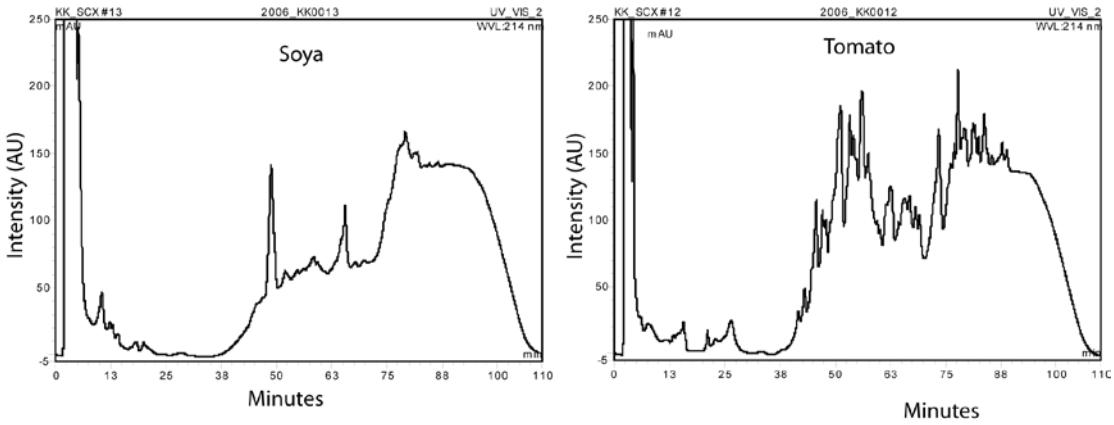


Figure 5.5 Offline cation exchange fractions of MudPIT tryptic peptides from A) soya and B) tomatoes

Table 5.3 Proteins from tomato identified by Pro ID.

	Tomato proteins identified with Pro ID (confidence >95%)		
Protein Database	SCX Cartridge	SCX	OFFGEL
viridiplantae	139	183	90
nrdb	141	184	94

RP nanoLC ESI-qTOF MS/MS method and capillary LC methods for ESI-MS^n (ion trap) were developed and optimised. The online MudPIT system was not developed, since it is known that off-line methods work better for the application of iTRAQ. It is recommended by the manufacturers of iTRAQ to perform separations off-line. The online MudPIT system was not developed, since it is known that off-line methods work better for the application of iTRAQ.

5.6 Assessment of MudPIT to determine the most effective

protocol.

Figure 5.6 displays the work flow to give the most effective protocol for MudPIT.

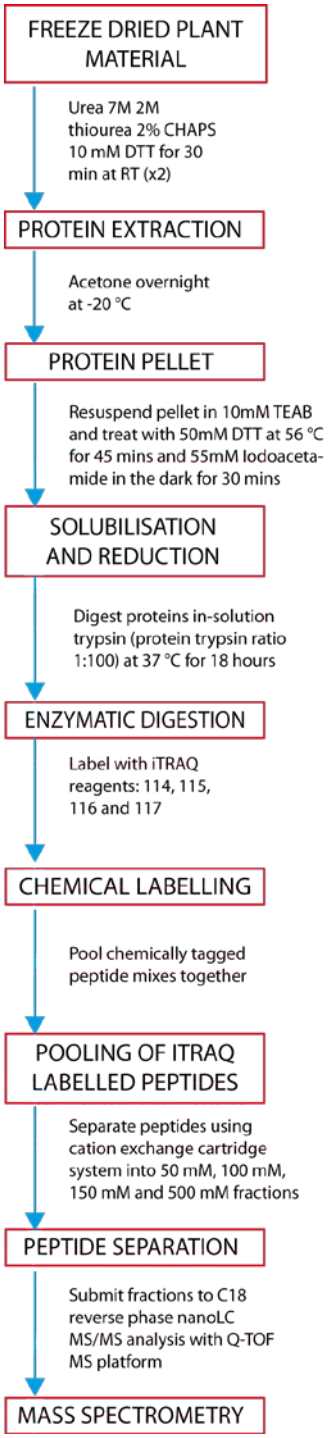


Figure 5.6 displays the work flow to give the most effective combination of procedure for MudPIT

5.7 Development of a quantitative MudPIT system for the

comparison of proteins in GM crops.

5.7.1 A “chemistry” method for quantitative MudPIT - iTRAQ

Soya and tomato proteins (100, 50 and 25 µg) were labelled with iTRAQ reagent and submitted to different MS platforms to identify differences in protein identification and quantification. The Q STAR (ABI) Q-TOF, Ion trap LCQ Deca and the Orbitrap (Thermo) were used with these samples. GM Soya was labelled with iTRAQ reagent and tested on the MALDI-TOF TOF MS platform. Tomato samples from GM and non-GM tomatoes were also tested on the Agilent 6520 Q-TOF (Agilent).

5.7.1.1 Compatibility of the iTRAQ method with the Ion TRAP (LCQ Deca)

iTRAQ labelled peptides from tomato samples were analysed with both the ion trap and Q-TOF instruments. No reporter ions were visible in the MS/MS spectrum (Figure 5.7) from the LCQ Deca ion trap due to the low mass cut off indicating that the ion trap is not compatible with iTRAQ experiments. As the low mass cut off prevents quantification of peptides using iTRAQ, quantification was performed in MS^3 spectrum. iTRAQ reporter ions were detected in some peptides. Figure 5.8 displays MS^3 spectra. However, quantification could not be performed due to missing signals from peptides and lack of sensitivity at MS^3.

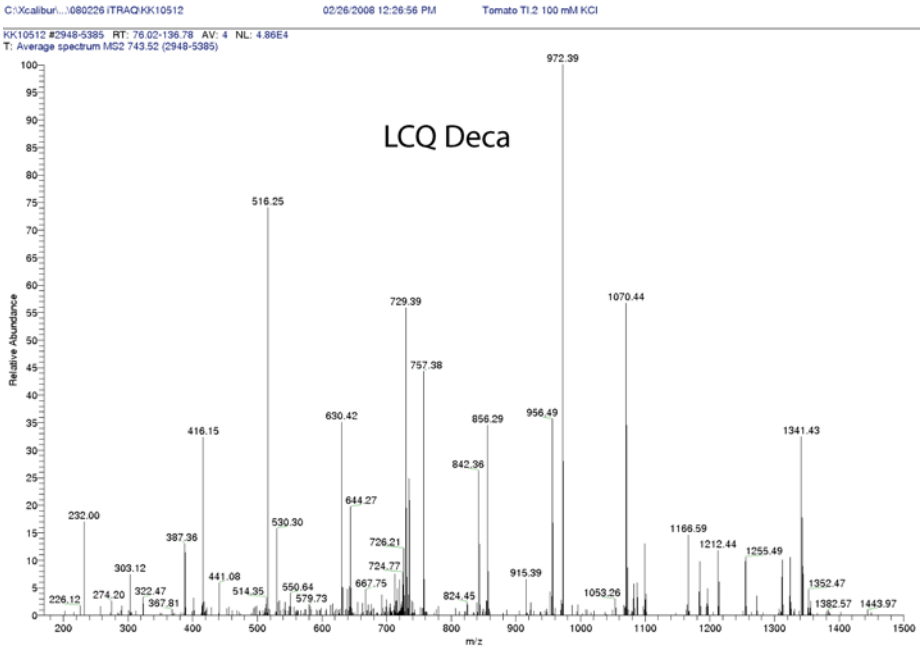


Figure 5.7 Product ion mass spectrum of iTRAQ labelled peptide SAINNLVNELVR from tomato sample analysed with both ion-trap (LCQ Deca). No reporter ions were generated due to low mass cut off.

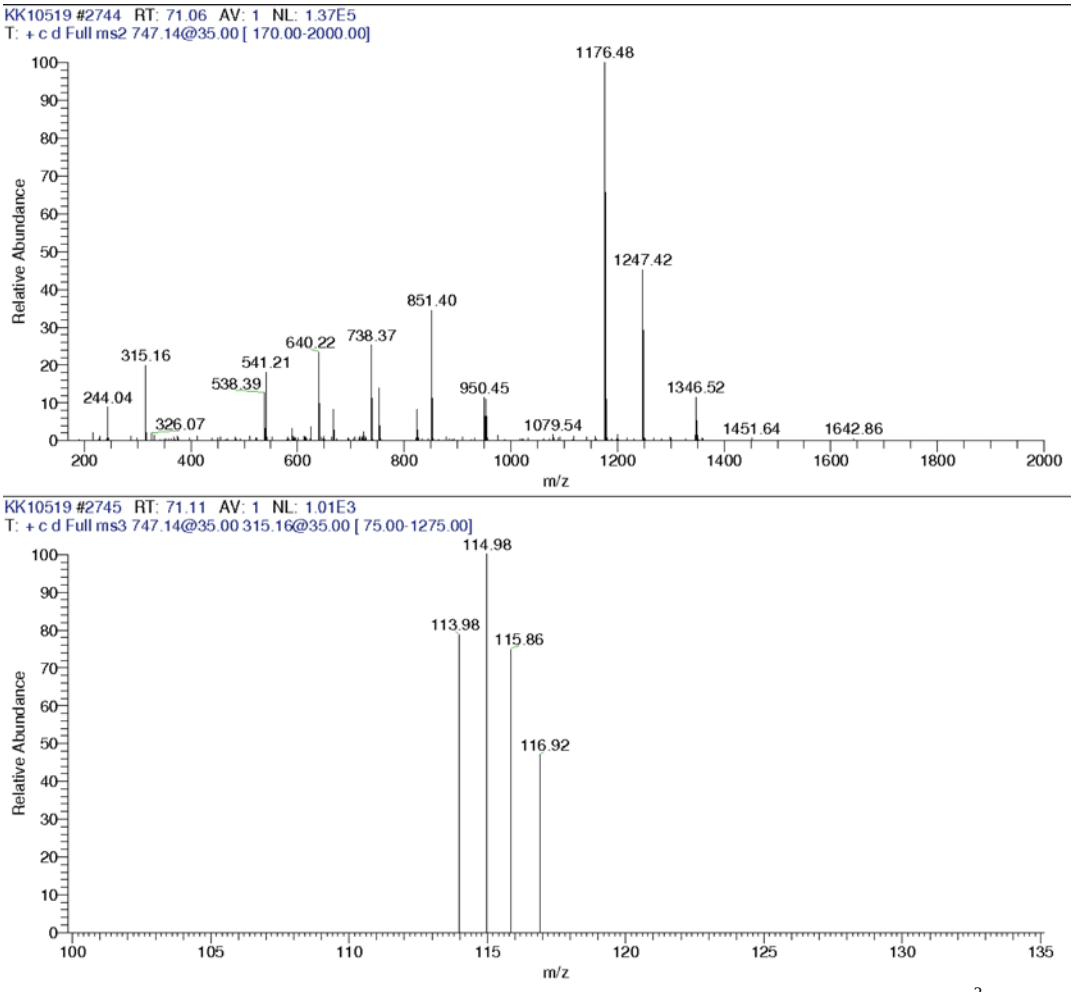


Figure 5.8 Product ion mass spectrum of iTRAQ labelled peptide and MS^3 spectrum from fragment 315 from tomato sample analysed with ion-trap (LCQ Deca).

5.7.1.2 iTRAQ as a 4-multiplex quantitative MudPIT approach using Orbitrap technology

iTRAQ labelled peptides from tomato and iTRAQ labelled peptides from soya were analysed using the ion trap with a FT MS platform. Figure 5.9 shows the TIC of cation exchange fractions collected from an iTRAQ soya experiment. The Orbitrap XL identified 374 proteins from tomato and 206 proteins from soya.

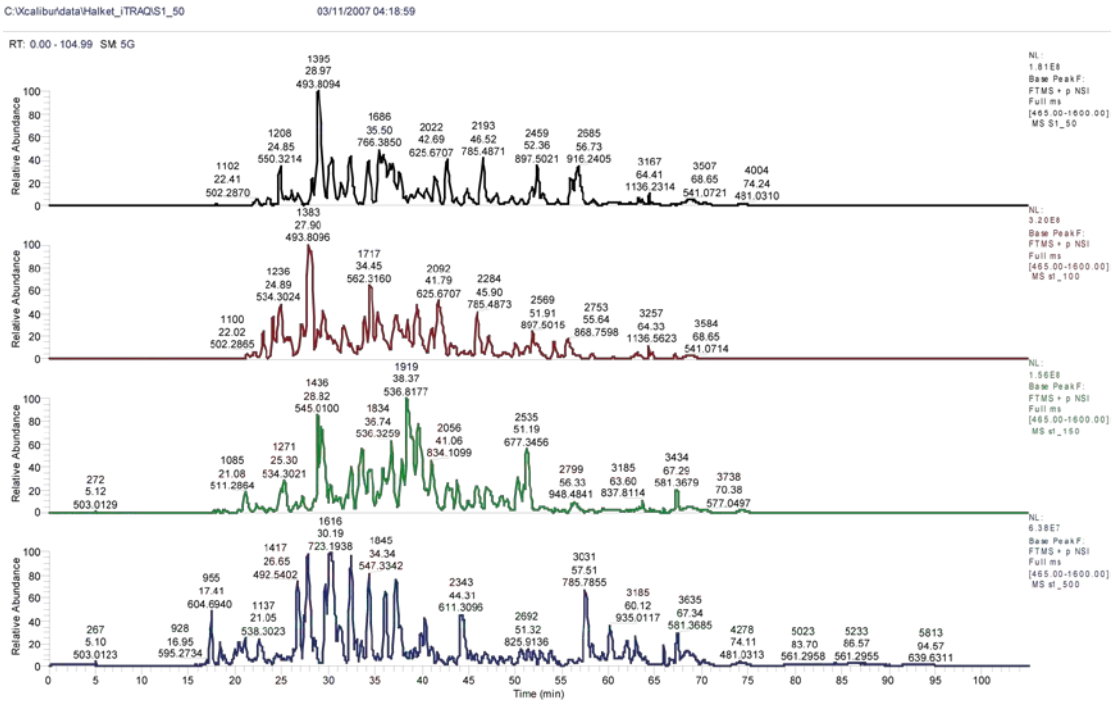


Figure 5.9 Product ion mass spectrum of iTRAQ labelled peptide and MS3 spectrum from fragment 315 from tomato sample analysed with ion-trap (Orbitrap).

5.7.1.3 iTRAQ as a 4-multiplex quantitative MudPIT approach using Q-TOF technology (Q-STAR (ABI) and Agilent 6520(Agilent))

iTRAQ labelled peptides from tomato were analysed using two Q-TOF MS platforms: The Q-STAR Pulsar i (ABI, Warrington) and the Agilent 6250 (Agilent). The QSTAR Pulsar i identified 91 proteins from tomato and

42 from soya. The Agilent 6520 Q TOF identified 350 proteins from tomato.

5.7.1.4 iTRAQ as a 4-multiplex quantitative MudPIT approach using MALDI-TOF TOF (Autoflex, Bruker Daltronics)

An iTRAQ method was developed for a non GM soya and GM soya (100 %) using a MALDI-TOF TOF MS platform. Non GM soya was labelled with 114 and GM soya with 115 and 116 reporter tags. Figure 5.10 gives a visual display of peptides identified in each cation exchange fraction. The MALDI TOF/TOF platform identified 157 proteins from soya.

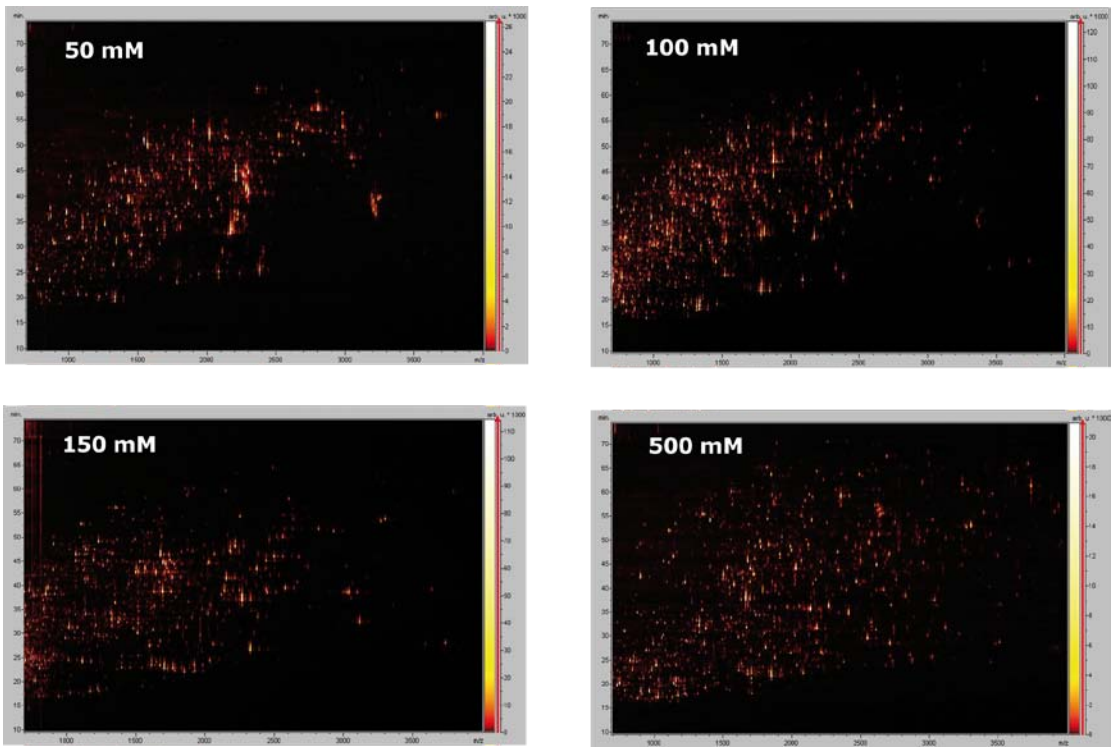


Figure 5.10 nanoLC-MALDI-MSMS analysis – LC-MALDI-MS heat map t R vs. m/z Peptide heat map representation of cation exchange fractions of iTRAQ labeled peptides using MALDI-TOF TOF.

5.7.1.5 iTRAQ - Identification of intended effects of genetic modification

In iTRAQ experiments performed on tomato and soya, respectively a wide selection of proteins were identified. These proteins include membrane bound proteins and proteins with extreme pIs. Using MudPIT the gene products derived from the transgene inserts were identified in both tomato (phytoene desaturase) and soya (EPSPS) samples by Q TOF MS (Figures 5.11 and Figures 5.12).

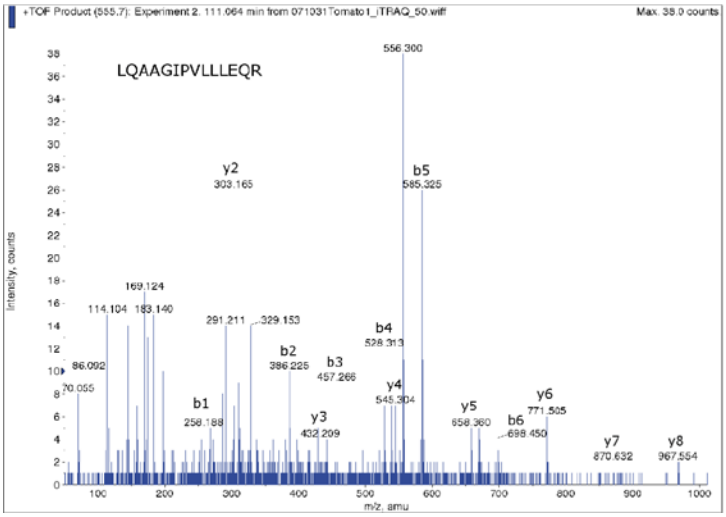
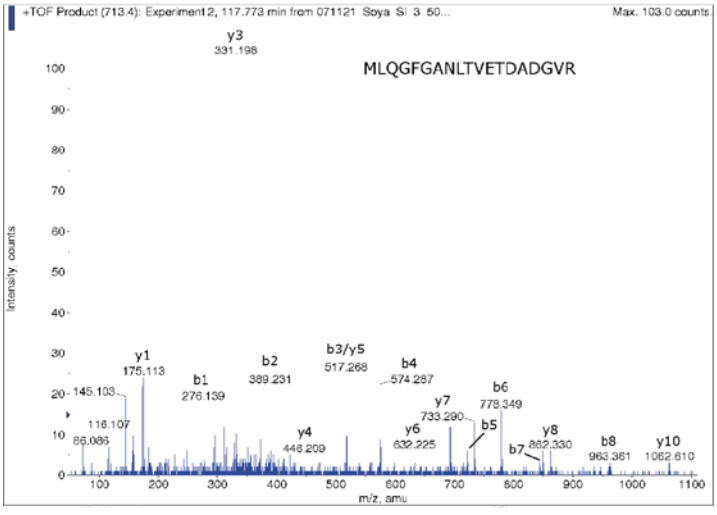


Figure 5.11 phytoene desaturase in GM CRT I tomatoes

	b-ions	y-ions
L	258.1934	1665.0180
Q	386.2520	1407.8318
A	457.2891	1279.7732
A	528.3262	1208.7361
G	585.3477	1137.6990
I	698.4317	1080.6776
P	795.4845	967.5935
V	894.5529	870.5407
L	1007.6370	771.4723
L	1120.7210	658.3883
L	1233.8051	545.3042
E	1362.8477	432.2201
Q	1490.9063	303.1775
R	1647.0074	175.1190



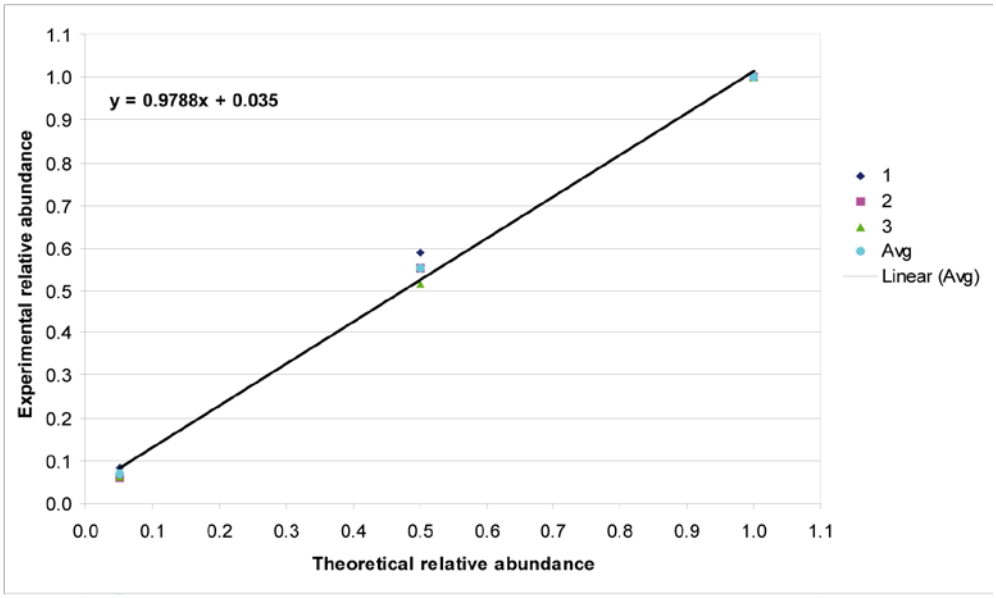
	b-ions	y-ions
M	276.1498	2138.0668
L	389.2339	1862.9243
Q	517.2925	1749.8402
G	574.3139	1621.7816
F	721.3823	1564.7602
G	778.4038	1417.6918
A	849.4409	1360.6703
N	963.4838	1289.6332
L	1076.5679	1175.5903
T	1177.6156	1062.5062
V	1276.6840	961.4585
E	1405.7266	862.3901
T	1506.7743	733.3475
D	1621.8012	632.2998
A	1692.8383	517.2729
D	1807.8653	446.2358
G	1864.8867	331.2088
V	1963.9552	274.1874
R	2120.0563	175.1190

Figure 5.12 5-enolpyruvylshikimate-3- phosphate synthase (EPSPS) in GM soya

5.7.1.6 Optimisation of the iTRAQ method and estimation of quantification of proteins

Four identical samples of tomato and of soya were labelled with 114, 115, 116 and 117 iTRAQ reagents and the set repeated three times. There were no major differences between the labels or the sample sets prepared on different days, indicating that the labelling procedure is reproducible.

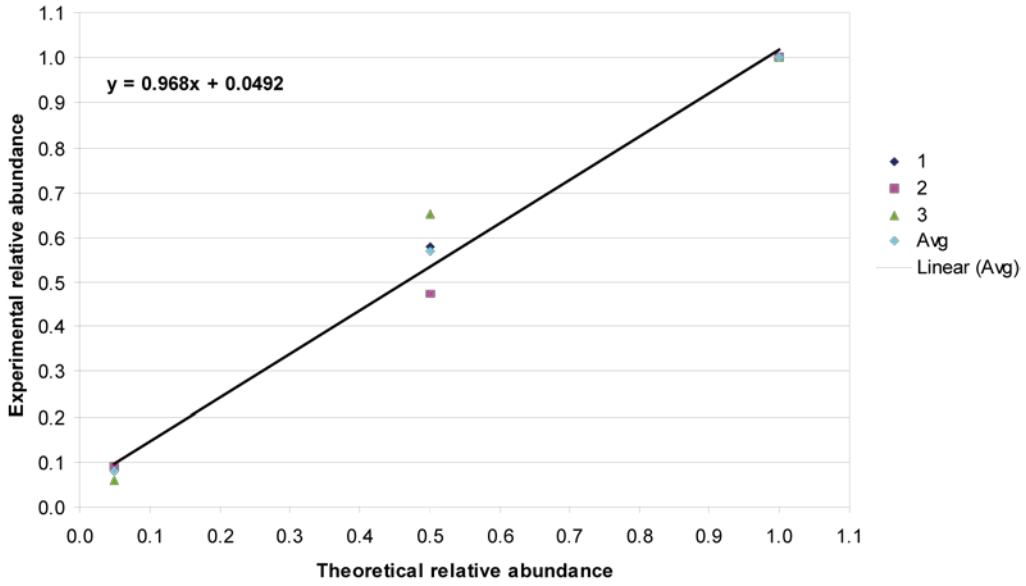
A peptide sample from GM soya was diluted 2-fold, 20-fold and 100-fold and labelled with different iTRAQ labels. This experiment was repeated three times. The 100-fold dilution could not be detected. Dilutions of soya samples at 2 and 20-fold gave mean quantities of 0.52-0.59 and 0.055-0.084. Coefficients of variance calculated for 30 proteins 10.2 % and 51.0%, respectively. In experiments, quantification of proteins showed the ratio to be slightly higher than the theoretical value (Figure 5.13).



Soya S II (30 proteins)	2-fold	20-fold	100-fold
Average ratios between 3 experiments	0.52-0.59	0.055-0.084	-
Average coefficient of variance (%)	10.2	51.0	-

Figure 5.13 A complex peptide sample from GM soya was diluted 2-fold, 20-fold and 100-fold and labeled with 114.1, 115.1,116.1 and 117.1.

A peptide sample from GM tomato (*Psy1* overexpressor) was diluted 2-fold, 20-fold and 40-fold and labelled with different iTRAQ labels. The experiment was repeated three times. The 40-fold dilution was quantified. However, the coefficient of variance was 63.3. Dilutions of tomato samples 2 and 20-fold gave an average mean of 0.47-0.65 and 0.061-0.088. Coefficients of variance calculated for 25 proteins were 21.5 % and 36.0 % respectively. Quantification of proteins showed the ratio to be slightly higher than the theoretical value (Figure 5.14).



Tomato T II (25 proteins)	2-fold	20-fold	40-fold
Average ratios between 3 experiments	0.47-0.65	0.061-0.088	0.034-0.063
Average coefficient of variance (%)	21.5	36.0	63.3

Figure 5.14 A peptide sample from GM-tomato (psy-1 over expressor) was diluted 2-fold, 20-fold and 40-fold and labelled with 114.1, 115.1,116.1 and 117.1.

5.7.1.7 Comparison of software packages for use with iTRAQ

Two software packages ProQuant and Protein Pilot were compared to address if protein quantification predictions were consistent. In this experiment a tomato extract was chemically tagged with label 114 and a 2-fold dilution was labelled with 115. The same data files collected from Q-TOF data acquisition were submitted for analysis. Table 5.4 displays the comparison between ProQuant and Protein Pilot using iTRAQ data from the Q-TOF platform.

Table 5.4 A comparison between different analysis software A) ProQuant and B) Protein pilot using Q-TOF data.

		115:114 = 1:2 ratio (2-fold dilution)	
		A ProQuant	B Protein pilot
Protein	Ratio	**115:114**	**115:114**
Acid β-fructofuranosidase precursor		0.6630	0.5751
Alcohol dehydrogenase 2		0.7037	0.6243
Enolase		0.6352	0.6351
Pectinesterase-1 precursor		0.7610	0.6579
1-aminocyclopropane-1-carboxylate oxidase homolog		0.6641	0.6403
Aconitate hydratase, cytoplasmic (Citrate hydro-lyase)		0.6776	0.6109
Stromal 70 kDa heat shock-related protein, chloroplast		0.6551	0.5983
Lipoxygenase B		0.6407	0.6181
Polygalacturonase-2 precursor (PG) (PG-2A) (PG-2B)		0.6615	0.6223
ATP synthase subunit beta, mitochondrial precursor		0.6657	0.6298

Acid β-fructofuranosidase is a common, abundant protein found in tomato. Four peptides identified from this protein with a confidence over 99 % were selected and the iTRAQ ratios examined. Table 5.5 displays the iTRAQ ratios of acid β-fructofuranosidase.

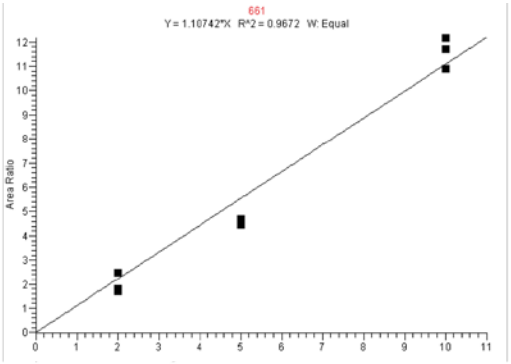
Table 5.5 A comparison between different peptides from acid β-fructofuranosidase

Protein	Confidence	Sequence	115:114
Acid β-fructofuranosidase precursor	99	ASLDDNKQDHYAIGTYDLGK	0.478
Acid β-fructofuranosidase precursor	99	GNPVLVPPPGIGVK	0.497
Acid β-fructofuranosidase precursor	99	LLVDHSIVESFAQGGR	0.641
Acid β-fructofuranosidase precursor	99	TGTHLLQWPVEEIESLR	0.718
		Mean	**0.583**

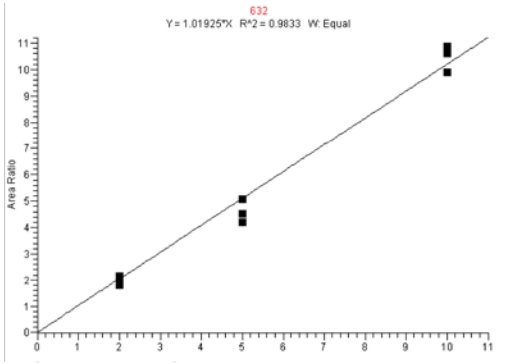
5.7.2 A "non-chemistry" method for quantitative MudPIT

Label free quantification with capillary LC MS/MS (LCQ Deca Ion Trap) system gave linear quantitation for soya peptide glycinin G2 precursor protein with a peptide internal standard (leucine enkephalin) (Figure 5.15)

There is potential to use label free quantitation with MudPIT samples for quantification of selected peptides. However, it would require a faster and more sensitive analysis platform e.g. Orbitrap.



Quantitation of peptide RPSYTNGPQEIYIQQGK
from Glycinin G1 precursor protein in GM soya sample
analysed with capillary LC-MS/MS (LCQ Deca ion trap).



Quantitation of peptide EAFGVNMQIVR
from Glycinin G2 precursor protein in GM soya sample
analysed with capillary LC-MS/MS (LCQ Deca ion trap).

x-axis: sample concentration (2 = 0.13 µg/µl, 5 = 0.33 µg/µl, 10 = 0.67 µg/µl)
y-axis: peptide:internal standard (leucine-enkephalin) ratio

Figure 5.15 A "non-chemistry" label-free method for quantitative MudPIT

5.7.3 Comparison of non-chemistry and chemistry based methods to quantitative proteomics

Label free quantification with capillary LC MS/MS (LCQ Deca Ion Trap) clearly gave better results than iTRAQ with the same instrument. The iTRAQ approach with nano LC MS/MS (QSTAR *Pulsar i*) was more sensitive and faster than label free. In iTRAQ the samples can be multiplexed into 4 or more recently 8.

5.8 A quantification MudPIT approach for the detection of intended and unintended effects resulting from genetic modification.

The MudPIT iTRAQ method with Agilent 6520 Q-TOF MS platform was chosen to assess the detection of intended and unintended effects resulting from genetic manipulation of tomato plants developed at Royal Holloway. The tomato cultivars being analysed in this section of study are listed in Table 5.6.

Table 5.6 Tomato cultivars grown to assess protein perturbations using MudPIT

Cultivar	GM modification	Phenotype
Ailsa craig	None	Wild type
Azygous	Been through the transformation process with no transgene incorporated	Identical to *Ailsa craig*
CRT I	Bacterial phytoene desaturase (Crt I)	Elevated levels of β-carotene
Psy1 Antisense	Down regulation of *psy1* in fruit	Lower levels of carotenoids
Psy1 Sense	Up regulation of *psy1* in fruit	Higher levels of carotenoids

Table 5.7 displays the iTRAQ labelled samples and each experiment undertaken. Experiments 1 - 16 were used to analyse the differences existing between plants in each cultivar line. Experiments 1 - 4 assessed the variation between plants in *Ailsa Craig*, Experiments 5 - 8 the azygous cultivars, experiments 9 - 12 the *Psy1* Antisense cultivars and in experiments 13 -16 the *psy1* Sense cultivars. Experiments 17 - 20 were used to address the differences between cultivars *Ailsa Craig*, azygous, *Psy1* Antisense and P*sy1* sense.

Table 5.7 iTRAQ experiments to assess the differences between plants of different cultivars and to assess the difference between cultivars.

	iTRAQ label			
Experiment	114	115	116	117
1	AC1	AC2	AC3	AC4
2	AC2	AC3	AC4	AC1
3	AC3	AC4	AC1	AC2
4	AC4	AC1	AC2	AC3
5	AZ1	AZ2	AZ3	AZ4
6	AZ2	AZ3	AZ4	AZ1
7	AZ3	AZ4	AZ1	AZ2
8	AZ4	AZ1	AZ2	AZ3
9	*Psy1* AS1	*Psy1* AS2	*Psy1* AS3	*Psy1* AS4
10	*Psy1* AS2	*Psy1* AS3	*Psy1* AS4	*Psy1* AS1
11	*Psy1* AS3	*Psy1* AS4	*Psy1* AS1	*Psy1* AS2
12	*Psy1* AS4	*Psy1* AS1	*Psy1* AS2	*Psy1* AS3
13	*Psy1* S1	*Psy1* S2	*Psy1* S3	*Psy1* S4
14	*Psy1* S2	*Psy1* S3	*Psy1* S4	*Psy1* S1
15	*Psy1* S3	*Psy1* S4	*Psy1* S1	*Psy1* S2
16	*Psy1* S4	*Psy1* S1	*Psy1* S2	*Psy1* S3
17	AC	AZ	*Psy1* AS	*Psy*1 S
18	AZ	*Psy1* AS	*Psy1* S	AC
19	*Psy1* AS	*Psy1* S	AC	AZ
20	*Psy1* S	AC	AZ	*Psy1* AS

5.8.1 Using MudPIT to assess variability between plants of the same cultivar

Experiments 1 - 16 were undertaken to assess the variation between plants of the same cultivar. Experimental variation associated with reproducibility of the chromatography and MS system and technical variation between repetitions of the same experiment were additionally assessed.

Three fruit were pooled from each biological replicate from the tomato lines. Biological replicates from each line were labelled with iTRAQ reagents 114-117 and multiplexed to compare and assess the natural biological variation inherent to each tomato cultivar. The experiment was repeated four times, to assess technical variation.

5.8.1.1 Variability between plants - ANOVA of the number of proteins and peptides identified

Figure 5.16 display the number of proteins identified for each cultivar, whilst Figure 5.17 displays the number of peptides identified from each iTRAQ experiment. In order to assess the reproducibility of the liquid chromatography and MS system, samples were injected twice. The number of proteins identified by each injection were not significantly different using ANOVA. This shows that each injection is comparable and the method is robust. Figure 5.18 shows box whisker plots and ANOVA values for comparison between and within cultivars for the number of proteins identified. ANOVA was performed on plants within the same cultivar. There were no significant differences between plants of the same cultivar F =

>0.05 (AC=303; AZ = 0.0052; *Psy1* AS = 0.834; *Psy1* S = 0.03). One way ANOVA was performed comparing AC vs. AZ; *Psy1* AS and *Psy1* S, all values were significant and less than F= 0.01.

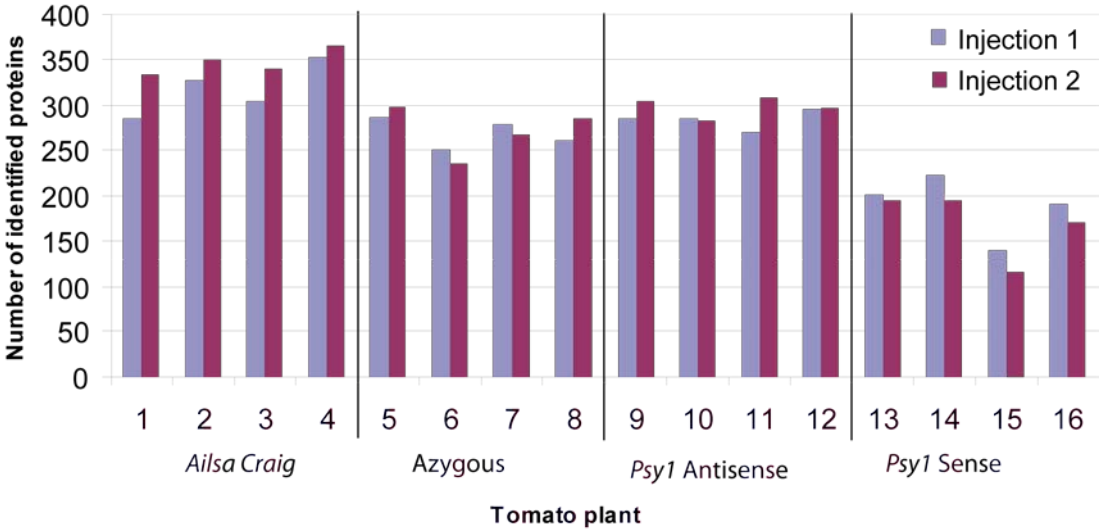


Figure 5.16 Bar charts displaying the number of proteins characterized from two injections into the mass spectrometry system.

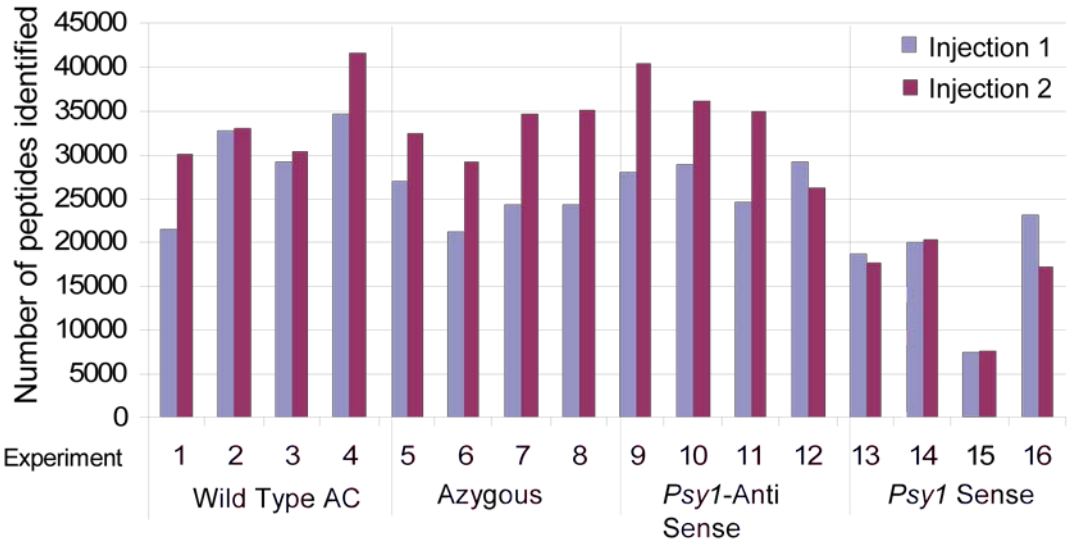
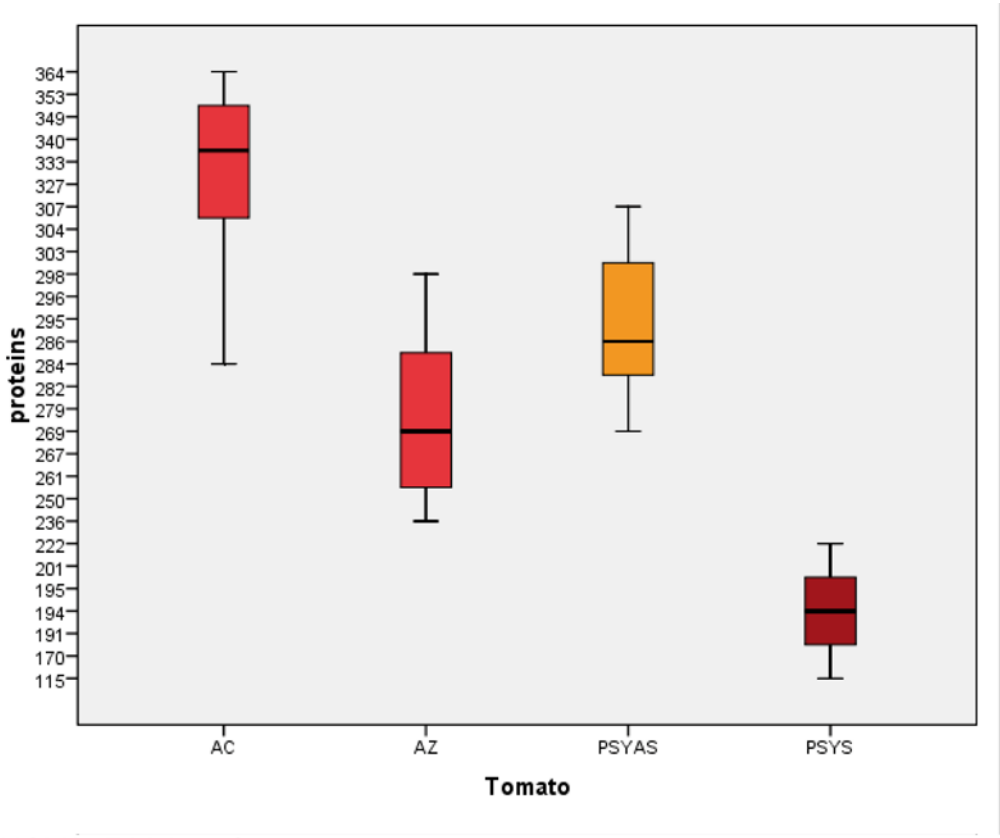


Figure 5.17 Bar charts displaying the number of peptides characterized from two injections into the mass spectrometry system.



ANOVA F Values for

Tomato	*Ailsa craig*	Azygous	*PSY1*-AS
Azygous	<0.00	-	
PSY-1-AS	0.01	0.035	-
PSY-1-S	<0.00	< 0.00	<0.00

Figure 5.18 Box whisker plots displaying the ANOVA of proteins characterized in the iTRAQ experiments

Figure 5.19 displays iTRAQ ratios plotted against the intensity estimation from MASCOT 2.3 for each cultivar. This shows that the most variation seen in the model is in proteins with very low abundance. Proteins with high intensity score have all four iTRAQ values clustered together.

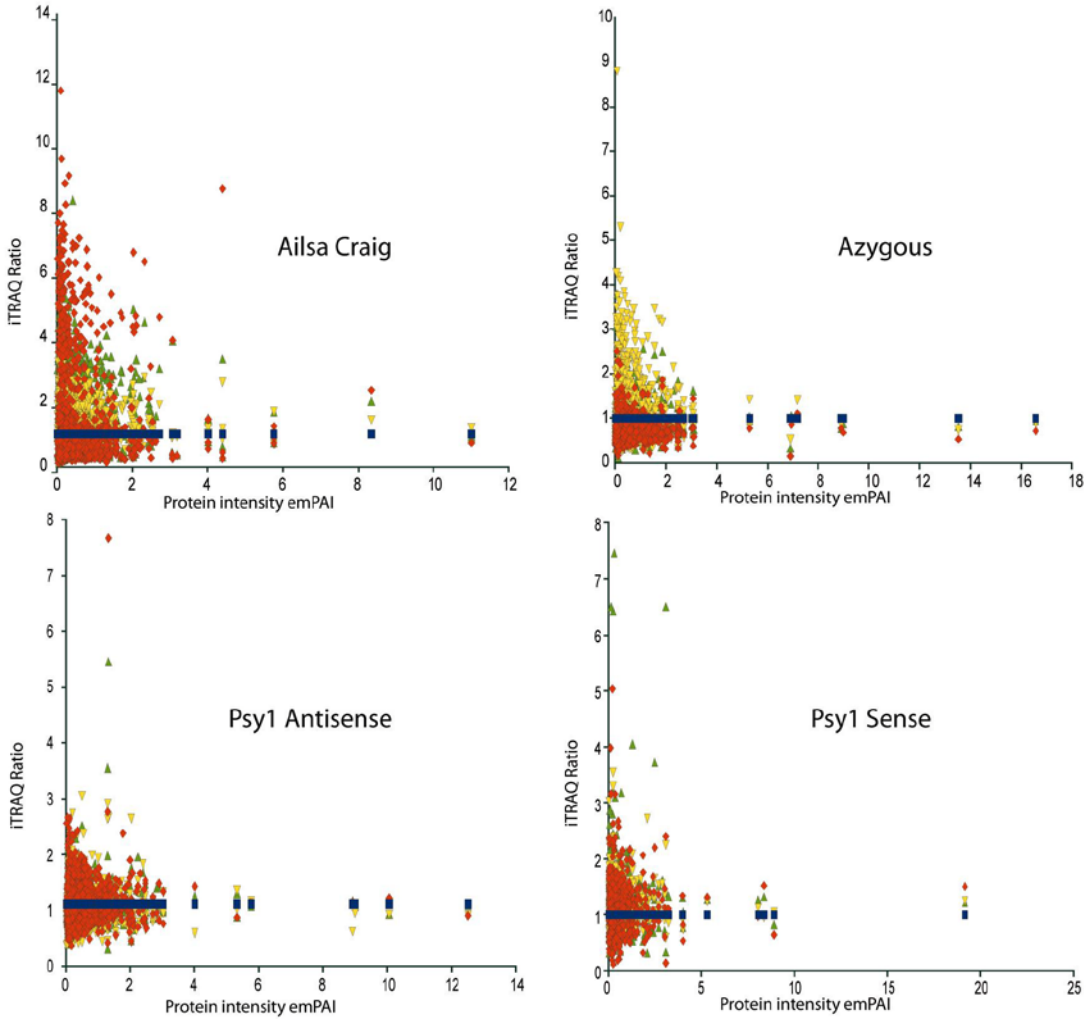


Figure 5.19 iTRAQ ratios plotted against emPAI intensity estimation from Mascot 3.2 Blue, 114/1114; red 115/114; orange 116/114; green 117/114.

5.8.1.2 Variability between plants and function of proteins identified in each cultivar

Figure 5.20 displays bar charts of the function of different types of protein identified in each cultivar. Proteins were classified into types: translation, structural, stress, seed storage, ribosomal, regulation, folding, enzymes, unknown and others. Stacked bar charts display the % composition, for ease of comparison between cultivars. In *Psy1* Sense, up to 10 % less enzymes were identified than in the wild type. Wild type, azygous, *Psy1* Antisense showed comparable percentage coverage of the different

classes of proteins, with between 95 - 130 enzymes identified in the MudPIT experiment.

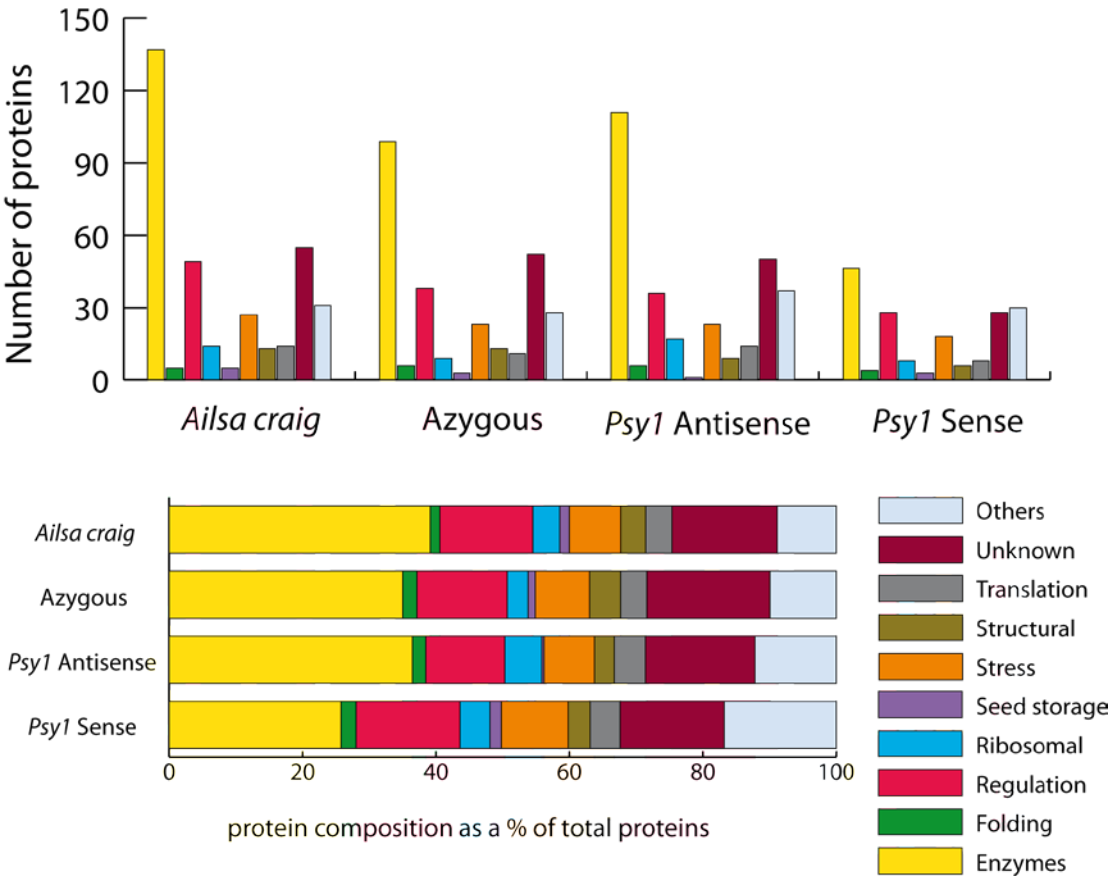


Figure 5.20 Bar charts displaying the number of proteins characterized from each cultivar and stacked bar charts showing the % composition of proteins identified

5.8.2 Using MudPIT iTRAQ to assess variability in protein composition between cultivars

5.8.2.1 Variability between cultivars - Students *t*-test

Table 5.8 gives the results of the Students *t*-test. Protein data was compiled into a data matrix using macros designed in Excel. The *t*-tests show individual fluctuations in specific levels of proteins when compared to *Ailsa Craig*. For the azygous cultivar 7 proteins were found to be significantly different from the wild type: 1-aminocyclopropane-1-carboxylate oxidase 1,

dihydrolipoyllysine-residue succinyl transferase component of 2-oxoglutarate dehydrogenase complex, disulfide-isomerase precursor-like protein, peptide methionine sulfoxide reductase, plastid isopentenyl diphosphate isomerase, vacuolar H+-ATPase A1 subunit isoform, osmotin I=group 5 basic pathogenesis-related protein were found to be significant to $P > 0.05$.

For *Psy1* Antisense, 11 proteins showed significant differences from the wild type. Enzymes identified include: ATP synthase subunit beta mitochondrial, peptidyl-prolyl *cis-trans* isomerase putative, phospholipid hydroperoxide glutathione peroxidase ***. House keeping proteins for regulation and structural functions included - 14-3-3-like protein D, calmodulin, elongation factor EF-2, Em-like protein, oxygen-evolving enhancer protein 2, chloroplastic **, remorin 1, small heat shock protein. *Psy1* Sense showed 67 significant differences, with 40 enzymes changed and 27 house keeping or structural proteins with significant differences to the wild type cultivar. Abscisic stress-ripening protein 1 was the only protein to be elevated against the control, other proteins were down regulated. Table 5.9 details the function of proteins identified as statistically different to the wild type cultivar by the *t*-test. The mean ratio is marked along with an average fold change.

Table 5.8 Tomato fruit at seven days post breaker have been analyzed. Protein levels are expressed as a ratio to *Ailsa Craig*. Three fruit from a minimum of four plants were used. The fruit were pooled and at least three determinations made per sample. The mean ratio data are presented ± SD. Students *t*-tests were used to determine significant differences between respective wild type backgrounds and transgenic varieties. $P < 0.05$, $P<0.01$ and $P < 0.001$ are indicated by *, **, *** respectively. Values in bold indicate where significant differences have been found compared to the wild type backgrounds. ND= not detected.

	Tomato genotype										
Protein ratio	Azygous				Psy1 Antisense				Psy1 Sense		
	Mean		SD	t-test	Mean		SD	t-test	Mean		SD	t-test
1-aminocyclopropane-1-carboxylate oxidase	1.27	±	0.82		1.16	±	0.75		**0.64**	**±**	**0.19**	*
1-aminocyclopropane-1-carboxylate oxidase 1	**0.42**	**±**	**0.25**	*	1.53	±	0.59		**0.49**	**±**	**0.3**	*
2-isopropylmalate synthase A	1.03	±	0.95		0.96	±	0.58		**0.48**	**±**	**0.16**	*
3-isopropylmalate dehydratase	0.86	±	0.45		1.42	±	0.61		0.57	±	0.33	
3-ketoacyl CoA thiolase 2	0.95	±	0.99		ND	±	ND		0.7	±	0.36	
Acid beta-fructofuranosidase	0.85	±	0.72		1.11	±	0.54		**0.57**	**±**	**0.24**	*
adenine phosphoribosyltransferase-like	0.63	±	0.23		1.2	±	0.6		0.46	±	0.24	
alcohol acyl transferase	0.84	±	1.09		0.95	±	0.63		**0.32**	**±**	**0.22**	*
alcohol dehydrogenase	1.31	±	1.61		ND	±	ND		ND	±	ND	
Alcohol dehydrogenase 2	0.78	±	0.58		0.98	±	0.62		**0.42**	**±**	**0.16**	**
alcohol dehydrogenase class III-like protein	0.35	±	0.3		0.97	±	0.56		**0.39**	**±**	**0.09**	**
aminoaldehyde dehydrogenase 2	0.52	±	0.48		1.05	±	0.72		0.5	±	0.38	
aspartic protease	0.82	±	0.66		0.99	±	0.68		**0.52**	**±**	**0.27**	*
ATP synthase CF1 alpha chain	0.77	±	0.89		1.05	±	1.13		**0.31**	**±**	**0.17**	*
ATP synthase CF1 beta chain	1.28	±	1.1		1.22	±	0.73		0.44	±	0.24	
ATP synthase CF1 epsilon chain	0.72	±	0.37		1.04	±	0.54		**0.53**	**±**	**0.16**	**
ATP synthase subunit alpha, chloroplastic	ND	±	ND		ND	±	ND		0.4	±	0.27	
ATPase-like	0.62	±	0.73		ND	±	ND		ND	±	ND	
ATP synthase subunit beta, mitochondrial	ND	±	ND		**0.36**	**±**	**0.09**	*	0.38	±	0.26	
carbonic anhydrase	0.67	±	0.52		1.02	±	0.91		**0.47**	**±**	**0.27**	*
Catalase isozyme 1	0.82	±	0.89		0.89	±	0.57		**0.48**	**±**	**0.32**	*
Complex Between Pectin Methylesterase And Its Inhibitor Protein	1.14	±	1.49		0.68	±	0.57		0.35	±	0.26	
cytosolic aconitase	0.91	±	0.86		1.05	±	0.47		**0.55**	**±**	**0.25**	*
cytosolic ascorbate peroxidase 1	0.95	±	0.67		0.99	±	0.32		0.78	±	0.3	
cytosolic nucleoside diphosphate kinase	0.77	±	0.46		0.83	±	0.36		**0.67**	**±**	**0.22**	*
dehydroascorbate reductase	0.86	±	0.53		0.94	±	0.35		**0.66**	**±**	**0.23**	*
dehydroascorbate reductase	0.93	±	0.1		1.06	±	0.57		0.8	±	0.42	
dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	**0.24**	**±**	**0.11**	*	0.59	±	0.45		**0.38**	**±**	**0.1**	*
disulfide-isomerase precursor-like protein	**0.47**	**±**	**0.19**	*	0.94	±	0.38		**0.65**	**±**	**0.1**	*
endopolygalacturonase	0.87	±	0.9		0.86	±	0.68		0.62	±	0.16	
Enolase	0.82	±	0.76		0.98	±	0.47		**0.49**	**±**	**0.3**	*
Ethylene-responsive proteinase inhibitor 1	0.83	±	0.55		1.23	±	0.73		1.4	±	1.24	
Glucan endo-1,3-beta-glucosidase B	0.36	±	0.3		1.28	±	1.35		0.77	±	0.65	
glucuronosyl transferase homolog, ripening-related	0.82	±	0.89		1.12	±	0.8		**0.37**	**±**	**0.25**	*
Glutaredoxin	0.74	±	0.39		0.87	±	0.17		**0.82**	**±**	**0.09**	*
Glutathione S-transfera	0.85	±	0.11		1.04	±	0.44		0.57	±	0.36	
Glyceraldehyde-3-phosphate dehydrogenase	0.89	±	0.98		0.9	±	0.61		**0.52**	**±**	**0.29**	*
ketol-acid reductoisomerase	1.02	±	0.4		1.26	±	0.8		**0.51**	**±**	**0.11**	**
Leucine aminopeptidase 2, chloroplastic	1.21	±	0.72		1	±	0.57		**0.51**	**±**	**0.21**	**
lipoxygenase	0.59	±	0.21		ND	±	ND		0.45	±	0.21	
lipoxygenase	0.47	±	0.44		0.95	±	0.52		**0.42**	**±**	**0.24**	**
lipoxygenase	1.07	±	1.2		0.86	±	0.43		**0.43**	**±**	**0.26**	*
Lipoxygenase A	0.79	±	0.89		0.85	±	0.38		**0.4**	**±**	**0.25**	*

malate dehydrogenase	0.96	±	0.89		1.14	±	0.52		**0.52**	±	**0.28**	*
methionine synthase	0.86	±	1.02		1.01	±	0.59		**0.4**	±	**0.27**	*
mitochondrial malate dehydrogenase	0.88	±	0.74		0.85	±	0.66		**0.46**	±	**0.23**	**
Monodehydroascorbate reductase	0.33	±	0.29		1.08	±	1.04		**0.27**	±	**0.08**	*
NADH-ubiquinone oxidoreductase 18 kDa subunit	0.8	±	0.88		0.83	±	0.29		0.77	±	0.35	
Pectinesterase/pectinesterase inhibitor U1	0.85	±	0.93		1.02	±	1.05		**0.25**	±	**0.1**	*
Peptide methionine sulfoxide reductase	**0.45**	±	**0.28**	*	0.9	±	0.42		0.68	±	0.27	
Peptidyl-prolyl cis-trans isomerase	0.97	±	0.69		1.04	±	0.44		0.67	±	0.34	
peptidyl-prolyl cis-trans isomerase, putative	0.49	±	0.47		**0.46**	±	**0.05**	*	**0.39**	±	**0.07**	**
phosphoenolpyruvate carboxylase 2	0.84	±	0.69		1.01	±	0.5		**0.53**	±	**0.26**	*
phosphoglycerate kinase precursor	1.12	±	1.12		0.84	±	0.46		0.54	±	0.21	
Phosphoglucomutase, cytoplasmic	0.3	±	0.31		0.98	±	0.96		**0.24**	±	**0.06**	*
phospholipid hydroperoxide glutathione peroxidase	0.57	±	0.54		**0.62**	±	**0**	***	0.87	±	0.26	
plastid isopentenyl diphosphate isomerase	**0.47**	±	**0.32**	*	1.02	±	0.49		0.36	±	0.17	
Polygalacturonase-2	0.33	±	0.33		0.85	±	0.63		**0.45**	±	**0.26**	*
protein phosphatase 2C	0.96	±	0.8		0.84	±	0.65		**0.65**	±	**0.26**	*
pyrophosphate-fructose 6-phosphate 1-phosphotransferase beta-subunit	0.29	±	0.27		1.08	±	0.96		**0.36**	±	**0.12**	*
S-adenosylmethionine synthetase 1	0.87	±	0.73		0.96	±	0.54		0.5	±	0.23	
Suberization-associated anionic peroxidase 1	0.97	±	0.57		0.98	±	0.19		0.87	±	0.37	
Superoxide dismutase	0.91	±	0.48		1.05	±	0.3		0.85	±	0.2	
Superoxide dismutase [Cu-Zn], chloroplastic	0.97	±	0.44		1.04	±	0.27		0.96	±	0.23	
thioredoxin peroxidase 1	0.82	±	0.52		0.92	±	0.31		**0.76**	±	**0.13**	*
UDP-glucose:protein transglucosylase-like	0.34	±	0.3		0.99	±	0.82		**0.28**	±	**0.04**	*
UTP--glucose-1-phosphate uridylyltransferase	0.4	±	0.33		0.85	±	0.91		0.31	±	0.2	
vacuolar H+-ATPase A1 subunit isoform	**0.18**	±	**0.05**	*	0.74	±	0.63		**0.2**	±	**0.06**	*
vicilin	1.6	±	1.03		0.73	±	0.72		**0.44**	±	**0.22**	**
Wound-induced proteinase inhibitor 1	1.02	±	0.59		1.23	±	0.38		**0.67**	±	**0.03**	***
House keepinhg and structural proteins										±		
14-3-3 protein	0.78	±	0.59		0.61	±	0.59		**0.55**	±	**0.25**	*
14-3-3-like protein D	0.49	±	0.54		**0.35**	±	**0.06**	*	0.47	±	0.28	
28 kDa ribonucleoprotein, chloroplastic	1.04	±	0.51		0.64	±	0.46		**0.68**	±	**0.11**	**
2S seed albumin-1 large subunit	10.87	±	16		9.1	±	1.13		0.9	±	0.9	
39 kDa EF-Hand containing protein	0.85	±	0.5		0.56	±	0.3		0.83	±	0.17	
40S ribosomal protein S20, putative	0.73	±	0.83		0.68	±	0.44		0.53	±	0.3	
40S ribosomal protein SA	0.38	±	0.35		0.41	±	1		**0.36**	±	**0.13**	*
60s acidic ribosomal protein-like protein	0.88	±	0.64		0.67	±	0.47		0.67	±	0.3	
60S ribosomal protein L13	0.73	±	0.42		0.48	±	0.66		**0.57**	±	**0.25**	*
Abscisic stress-ripening protein 1	1.15	±	0.07		0.51	±	0.38		**1.16**	±	**0.1**	*
ACBP6 (acyl-CoA-binding protein 6)	0.55	±	0.37		0.46	±	0.21		1.23	±	0.57	
acidic ribosomal protein P1a-like	0.73	±	0.48		0.47	±	0.26		**0.6**	±	**0.27**	*
actin	1.01	±	0.89		0.79	±	0.46		**0.52**	±	**0.27**	*
ADP,ATP carrier protein, mitochondrial	1.1	±	1.43		1.01	±	0.64		0.35	±	0.28	
ankyrin-repeat protein HBP1	0.64	±	0.69		0.46	±	0.1		0.64	±	0.38	
calmodulin	0.71	±	0.43		0.39	±	0.17	*	0.81	±	0.28	
chaperonin 21 precursor	0.96	±	0.28		0.53	±	0.42		**0.74**	±	**0.18**	*
chaperonin 60 alpha subunit	0.47	±	0.33		0.43	±	0.56		**0.58**	±	**0.22**	*
chaperonin-60 beta subunit	0.84	±	0.62		ND	±	ND		ND	±	ND	
class II small heat shock protein Le-HSP17.6	0.77	±	0.55		0.73	±	0.17		**0.53**	±	**0.17**	*
cold-stress inducible protein	0.94	±	0.39		0.48	±	0.26		1.09	±	0.2	
CP12	0.81	±	0.59		0.57	±	0.23		0.77	±	0.44	
cyclophilin	0.78	±	0.28		0.52	±	0.54		ND	±	ND	
Elongation factor 1-alpha	1.05	±	0.15		ND	±	ND		**0.7**	±	**0.01**	***
elongation factor EF-2	0.52	±	0.41		**0.52**	±	**0.11**	*	0.45	±	0.21	
Em protein	1.29	±	0.43		0.6	±	0.1		1.22	±	0.33	

Em-like protein	1.07	±	0.56		**0.32**	**±**	**0.2**	*	0.98	±	0.35	
eukaryotic translation initiation factor 4A	1.03	±	1.17		0.59	±	0.53		**0.62**	**±**	**0.32**	*
Eukaryotic translation initiation factor 5A-3	0.89	±	0.58		0.55	±	0.19		0.88	±	0.27	
fruit-ripening protein	0.49	±	0.42		0.83	±	0.23		0.79	±	0.05	
heat shock protein	0.77	±	0.64		0.54	±	0.67		**0.43**	**±**	**0.29**	*
heat shock protein 101	0.34	±	0.27		0.4	±	0.77		**0.33**	**±**	**0.07**	**
histone H2B-like protein	1.12	±	0.7		ND	±	ND		ND	±	ND	
histone H4	0.91	±	0.74		0.54	±	0.46		**0.53**	**±**	**0.22**	*
Hsc70	0.88	±	0.75		0.31	±	0.53		**0.54**	**±**	**0.3**	*
induced stolon tip protein NAP1Ps	0.84	±	0.19		0.91	±	0.45		0.69	±	0.12	
late embryogenesis (Lea)-like protein	0.84	±	0.58		0.67	±	0.52		**0.63**	**±**	**0.19**	*
LeArcA1 protein	0.33	±	0.25		ND	±	ND		ND	±	ND	
major intrinsic protein 2	0.87	±	1.07		0.63	±	0.7		0.29	±	0.21	
Mitochondrial outer membrane protein porin of 34 kDa	0.9	±	1		0.49	±	0.53		**0.48**	**±**	**0.23**	*
Mitochondrial-processing peptidase subunit alpha	0.41	±	0.4		ND	±	ND		ND	±	ND	
multiprotein bridging factor 1b	1.16	±	0.19		0.63	±	0.26		0.91	±	0.16	
osmotin l=group 5 basic pathogenesis-related protein	0.27	±	0.18	*	0.29	±	0.81		0.56	±	0.41	
Oxygen-evolving enhancer protein 1, chloroplastic	0.96	±	0.83		0.77	±	0.24		**0.63**	**±**	**0.17**	*
Oxygen-evolving enhancer protein 2, chloroplastic	0.77	±	0.38		**0.8**	**±**	**0.19**	**	0.76	±	0.37	
pathogenesis-related protein 10	0.74	±	0.5		0.4	±	0.43		0.69	±	0.22	
Pathogenesis-related protein P2	0.99	±	0.66		0.49	±	0.48		**0.59**	**±**	**0.2**	*
polyubiquitin	0.85	±	0.05			±			0.73	±	0.22	
Profilin-2	0.92	±	0.62		0.61	±	0.22		0.75	±	0.24	
proteasome, alpha subunit-like protein	0.59	±	0.47		0.39	±	0.46		0.62	±	0.35	
putative stress-induced protein	0.63	±	0.4		0.55	±	0.4		0.72	±	0.33	
Ran binding protein-1	0.71	±	0.44		0.71	±	0.42		**0.66**	**±**	**0.14**	*
remorin 1	0.85	±	0.45		**0.45**	**±**	**0.13**	*	**0.8**	**±**	**0.09**	*
ribosomal protein L12-1a	0.8	±	0.37		0.57	±	0.21		**0.64**	**±**	**0.19**	*
RNA-binding protein precursor	0.91	±	0.69		0.51	±	0.26		0.87	±	0.19	
small heat shock protein	0.9	±	0.56		**0.49**	**±**	**0.13**	*	**0.54**	**±**	**0.16**	**
Small heat shock protein, chloroplastic	0.81	±	0.45		0.46	±	0.22		0.46	±	0.26	
Small heat shock protein, chloroplastic	1.12	±	0.87		ND	±	ND		ND	±	ND	
translation elongation factor-1 alpha	0.84	±	0.74		0.44	±	0.6		**0.5**	**±**	**0.24**	*
Translationally-controlled tumor protein homolog	0.74	±	0.51		0.48	±	0.48		**0.62**	**±**	**0.26**	*

Table 5.9 Function, type and fold change of proteins with significant differences identified in azygous, *Psy1* Antisense and *Psy1* Sense cultivars.

Protein	Up/Down Regulated	Fold change	Type	Pathway	Function and Notes
Cultivar - Azygous					
1-aminocyclopropane-1-carboxylate oxidase 1	0.42	2.35 desrease	Enzymes	ethylene biosynthesis	Converts 1-aminocyclopropane-1-carboxylic acid (ACC) to ethylene, the last step in ethylene biosynthesis (Spanu et al., 1991)
disulfide-isomerase precursor-like protein	0.47	2.14 desrease	Enzyme	2-ketoglutarate dehydrogenase complex	Enzymatic reaction of: dihydrolipoyllysine-residue succinyltransferase
Peptide methionine sulfoxide reductase	0.45	2.24 desrease	Enzyme		
plastid isopentenyl diphosphate isomerase	0.47	2.15 desrease	Enzyme	isoprenoid biosynthesis	
vacuolar H+-ATPase A1 subunit isoform	0.18	5.71 desrease	Enzyme		Proton pump
dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	0.24	4.19 desrease	Enzyme	2-ketoglutarate dehydrogenase complex	
Cultivar - Psy1 Antisense					
calmodulin	0.7	1.43 desrease	regulation		Calcium regulation
elongation factor EF-2	0.74	1.34 desrease	translation facilitators		
remorin 1	0.7	1.42 desrease	regulation		
small heat shock protein	0.78	1.28 desrease	stress		
Small heat shock protein, chloroplastic	0.73	1.37 desrease	stress		
phospholipid hydroperoxide glutathione peroxidase	0.62	1.61 desrease	Enzyme	glutathione redox reactions I	Enzymatic reaction of: glutathione peroxidase
peptidyl-prolyl cis-trans isomerase, putative	0.46	2.18 desrease	Enzyme	Chaperone function in folding	PPIases accelerate the folding of proteins
Em protein	0.52	1.91 desrease	seed storage		
Em-like protein	0.42	2.37 desrease	seed storage		
Oxygen-evolving enhancer protein 2, chloroplastic	0.42	2.39 desrease	regulation		
14-3-3-like protein D	0.41	2.42 desrease	regulation		
Cultivar - Psy1 Sense					
Abscisic stress-ripening protein 1	1.16	1.19 increase	stress		
1-aminocyclopropane-1-carboxylate oxidase	0.64	1.56 desrease	Enzyme	ethylene biosynthesis	Converts 1 - aminocyclopropane - 1 - carboxylic acid (ACC) to ethylene
cytosolic nucleoside diphosphate kinase	0.67	1.49 desrease	Enzyme	leucine biosynthesis	
dehydroascorbate reductase	0.66	1.52 desrease	Enzyme		chloroplast enzymes involved in the protection against oxidative stress
disulfide-isomerase precursor-like protein	0.65	1.54 desrease	Enzyme		thioredoxin domains and controlling diversified metabolic functions, including disulfide bond formation and isomerisation during protein folding
Glutaredoxin	0.82	1.21 desrease	Enzyme		redox enzymes of approximately o which use glutathione as a cofactor

protein phosphatase 2C	0.65	1.53 desrease	Enzyme		removes a phosphate group from its substrate by hydrolysing phosphoric acid
thioredoxin peroxidase 1	0.76	1.32 desrease	Enzyme		act as antioxidants by facilitating the reduction of other proteins by cysteine thiol-disulfide exchange
Wound-induced proteinase inhibitor 1	0.67	1.48 desrease	regulation		
28 kDa ribonucleoprotein, chloroplastic	0.68	1.46 desrease	Ribosomal		
acidic ribosomal protein P1a-like	0.6	1.65 desrease	Ribosomal		
chaperonin 21 precursor	0.74	1.35 desrease	Folding		
Elongation factor 1-alpha	0.7	1.43 desrease	translation facilitators		
eukaryotic translation initiation factor 4A	0.62	1.61 desrease	translation facilitators		
late embryogenesis (Lea)-like protein	0.63	1.6 desrease	Stress		
Oxygen-evolving enhancer protein 1, chloroplastic	0.63	1.58 desrease	regulation		thioredoxins
Ran binding protein-1	0.66	1.51 desrease	inhibitor		guanine nucleotide dissociation inhibitory activity
remorin 1	0.8	1.26 desrease	regulation		
ribosomal protein L12-1a	0.64	1.57 desrease	Ribosomal		
Translationally-controlled tumor protein homolog	0.62	1.6 desrease	regulation		
1-aminocyclopropane-1-carboxylate oxidase 1	0.49	2.06 desrease	Enzyme	ethylene biosynthesis	Converts 1 - aminocyclopropane - 1 - carboxylic acid (ACC) to ethylene
2-isopropylmalate synthase A	0.48	2.1 desrease	Enzyme	leucine biosynthesis	
Acid beta-fructofuranosidase	0.57	1.75 desrease	Enzyme	sucrose degradation III	Enzymatic reaction of: β-fructofuranosidase (beta-fructofuranosidase)
Alcohol dehydrogenase 2	0.42	2.37 desrease	Enzyme	pyruvate fermentation to ethanol II	
aspartic protease	0.52	1.94 desrease	Enzyme		use an aspartate residue for catalysis of their peptide substrates
ATP synthase CF1 epsilon chain	0.53	1.9 desrease	Enzyme		Produces ATP from ADP in the presence of a proton gradient across the membrane
carbonic anhydrase	0.47	2.14 desrease	Enzyme	cyanate degradation	
Catalase isozyme 1	0.48	2.09 desrease	Enzyme	superoxide radicals degradation	
cytosolic aconitase	0.55	1.82 desrease	Enzyme	glyoxylate cycle , TCA cycle variation III (eukaryotic)	Enzymatic reaction of: 2-methylaconitate hydratase (aconitase//aconitate hydratase//aconitate hydratase)
Enolase	0.49	2.05 desrease	Enzyme	glycolysis IV (plant cytosol, glycolysis I	Enzymatic reaction of: ENOLASE
Glyceraldehyde-3-phosphate dehydrogenase	0.52	1.91 desrease	Enzyme	glycolysis IV (plant cytosol) , glycolysis I	Enzymatic reaction of: glyceraldehyde-3-phosphate dehydrogenase
ketol-acid reductoisomerase	0.51	1.97 desrease	Enzyme	isoleucine biosynthesis I (from threonine), valine biosynthesis	Enzymatic reaction of: KETOL-ACID REDUCTOISOMERASE
Leucine aminopeptidase 2, chloroplastic	0.51	1.95 desrease	Enzyme		Ribonuclease protection studies showed that LAP A2 has18bp short UTR region and is expressed after wounding

lipoxygenase	0.42	2.4 desrease	Enzyme	jasmonic acid biosynthesis	
lipoxygenase	0.43	2.32 desrease	Enzyme	jasmonic acid biosynthesis	
Lipoxygenase A	0.4	2.5 desrease	Enzyme	jasmonic acid biosynthesis	
malate dehydrogenase	0.52	1.91 desrease	Enzyme	glyoxylate cycle , gluconeogenesis I , TCA cycle variation III (eukaryotic)	
methionine synthase	0.4	2.47 desrease	Enzyme	methionine biosynthesis II , S-adenosyl-L-methionine cycle II	
mitochondrial malate dehydrogenase	0.46	2.15 desrease	Enzyme	pyruvate fermentation to lactate	
phosphoenolpyruvate carboxylase 2	0.53	1.89 desrease	Enzyme	gluconeogenesis I	
phosphoglycerate kinase precursor	0.54	1.85 desrease	Enzyme	formaldehyde assimilation III (dihydroxyacetone cycle)	
Polygalacturonase-2	0.45	2.23 desrease	Enzyme	homogalacturonan degradation	
S-adenosylmethionine synthetase 1	0.5	2.01 desrease	Enzyme		catalyzes the formation of S-adenosylmethionine (AdoMet) from methionine and ATP
vicilin	0.44	2.25 desrease	Seed storage		Potentially allergenic
14-3-3 protein	0.55	1.81 desrease	regulation		binds signaling proteins, including kinases, phosphatases, and transmembrane receptor
60S ribosomal protein L13	0.57	1.74 desrease	Ribosomal		
actin	0.52	1.92 desrease	Structural		
chaperonin-60 beta subunit	0.58	1.74 desrease	Folding		
class II small heat shock protein Le-HSP17.6	0.53	1.88 desrease	Stress		
heat shock protein	0.43	2.34 desrease	Stress		
histone H4	0.53	1.9 desrease	Structural		Package and order DNA into structural units called nucleosomes
Hsc70	0.54	1.85 desrease	Stress		
Mitochondrial outer membrane protein porin of 34 kDa	0.48	2.08 desrease	structual		
Pathogenesis-related protein P2	0.59	1.71 desrease	stress		
small heat shock protein	0.54	1.86 desrease	Stress		
translation elongation factor-1 alpha	0.5	2.01 desrease	translation facilitators		
alcohol dehydrogenase class III-like protein	0.39	2.57 desrease	Enzyme	pyruvate fermentation to ethanol II	
peptidyl-prolyl cis-trans isomerase, putative	0.39	2.57 desrease	Enzyme		PPIases accelerate the folding of proteins
dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	0.38	2.66 desrease	Enzyme		
glucuronosyl transferase homolog, ripening-related	0.37	2.67 desrease	Enzyme		
40S ribosomal protein SA	0.36	2.75 desrease	Ribosomal		
pyrophosphate-fructose 6-phosphate 1-phosphotransferase beta-subunit	0.36	2.79	Enzyme	glycolysis IV (plant cytosol) , glycolysis I	

heat shock protein 101	0.33	3.06 desrease	Stress		
alcohol acyl transferase	0.32	3.13 desrease	Enzyme	wax esters biosynthesis I	Enzymatic reaction of: fatty acyl-CoA reductase
UDP-glucose:protein transglucosylase-like	0.28	3.61 desrease	Enzyme		
Monodehydroascorbate reductase	0.27	3.77 desrease	Enzyme	ascorbate glutathione cycle	
Phosphoglucomutase, cytoplasmic	0.24	4.2 desrease	Enzyme		
Pectinesterase/pectinesterase inhibitor U1	0.25	4 desrease	Enzyme	homogalacturonan degradation	
vacuolar H+-ATPase A1 subunit isoform	0.2	5.09 desrease	Enzyme		

5.8.2.2 Variability between cultivars – enzymes and metabolites in sucrose degradation, glycolysis and TCA cycle pathways

Proteomic data from *Psy1* Sense was added to metabolomics data (Enfissi *et al.*, 2007) relating to sucrose degradation, glycolysis and TCA cycle. The pathway diagrams show elevated metabolites and enzymes in green, unchanged enzymes and metabolites in grey, metabolites and enzymes that show decreases in red and white to show that there is no value for an enzyme or metabolite. Figure 5.22 shows the enzymatic and metabolite data for sucrose degradation and glycolysis. Figure 5.23 shows visualised data for the TCA cycle and coupled reactions.

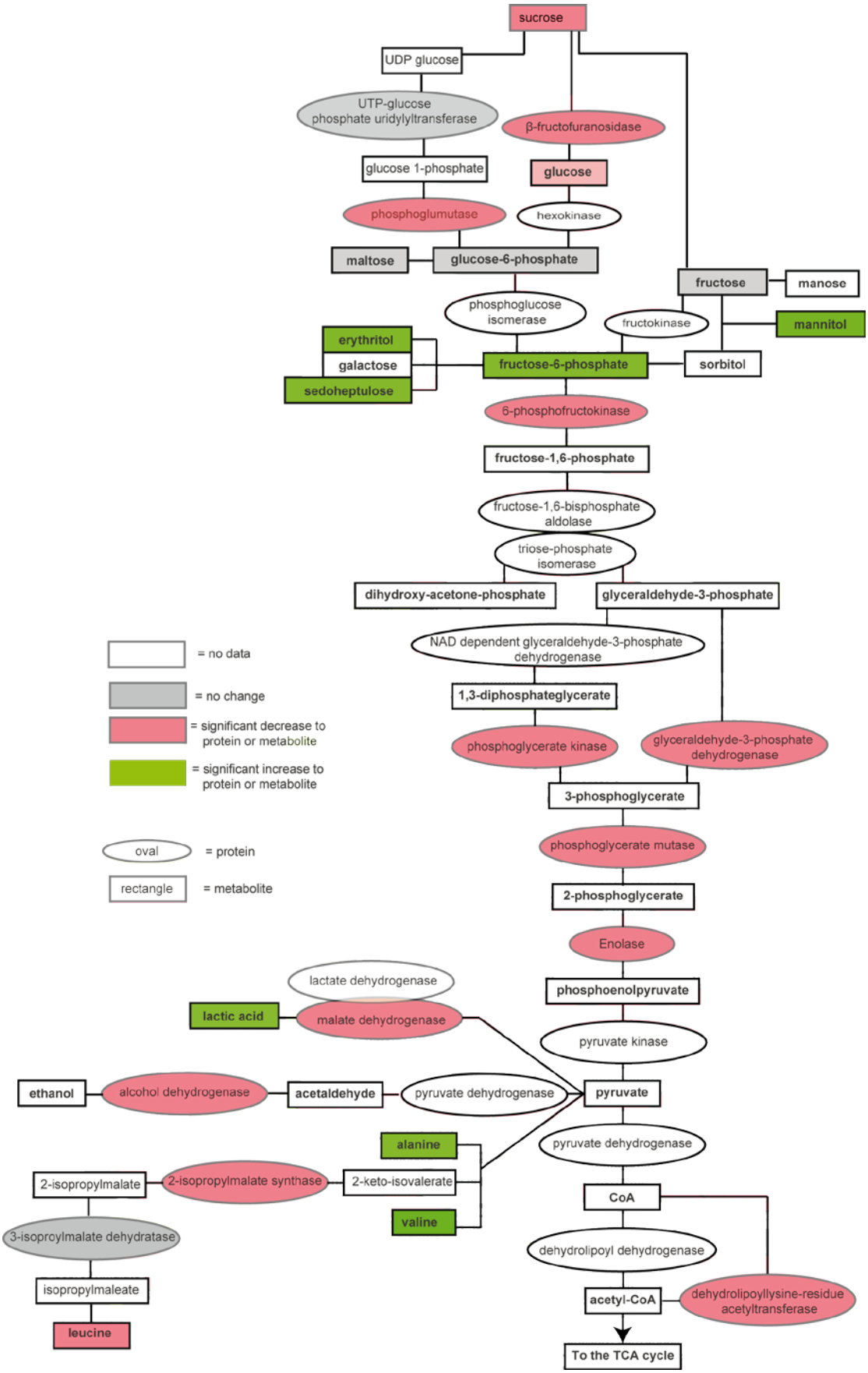


Figure 5.21 Protein levels in sucrose degradation and glycolysis pathways of Psy1 Sense red ripe fruit with metabolite data.

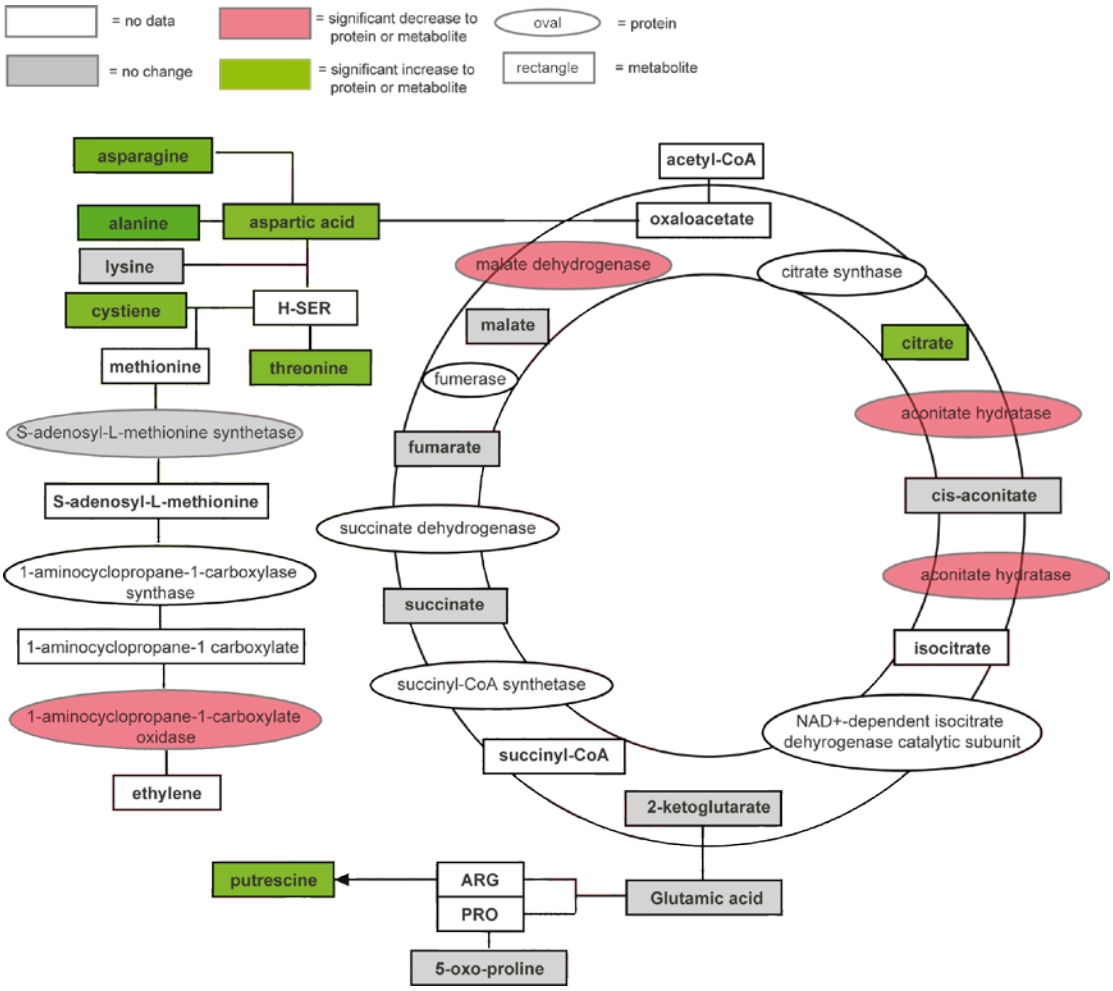


Figure 5.22 Protein levels in *Psy1* Sense cultivar with metabolite data

5.8.2.3 Variability between cultivars - Principal Component Analysis (PCA)

Data matrices were submitted for PCA and PLS-DA analysis. Figure 5.23 shows the PCA scatter plot and correlations loading plot for *Ailsa Craig (AC)*, azygous (AZ), *Psy1* Antisense and *Psy1* Sense tomato cultivars. From the PCA cluster plot no tomato cultivar formed a discrete grouping. There was not enough variation between protein levels in each the tomato cultivar to allow differentiation by PCA. AC, AZ, *Psy1* Antisense and *Psy1* Sense peptide profiles did not form discrete groups and were interspersed. This PCA diagram shows the limits of PCA, as there is no change between

samples. The loadings plot shows that some protein coefficients are dispersed and far away from the origin. These proteins show the most differences and account for the variability of the model. The protein at the base of the plot is abscisic stress-ripening protein 1. This protein was identified by the *t*-test as being significantly different in *Psy1* Sense.

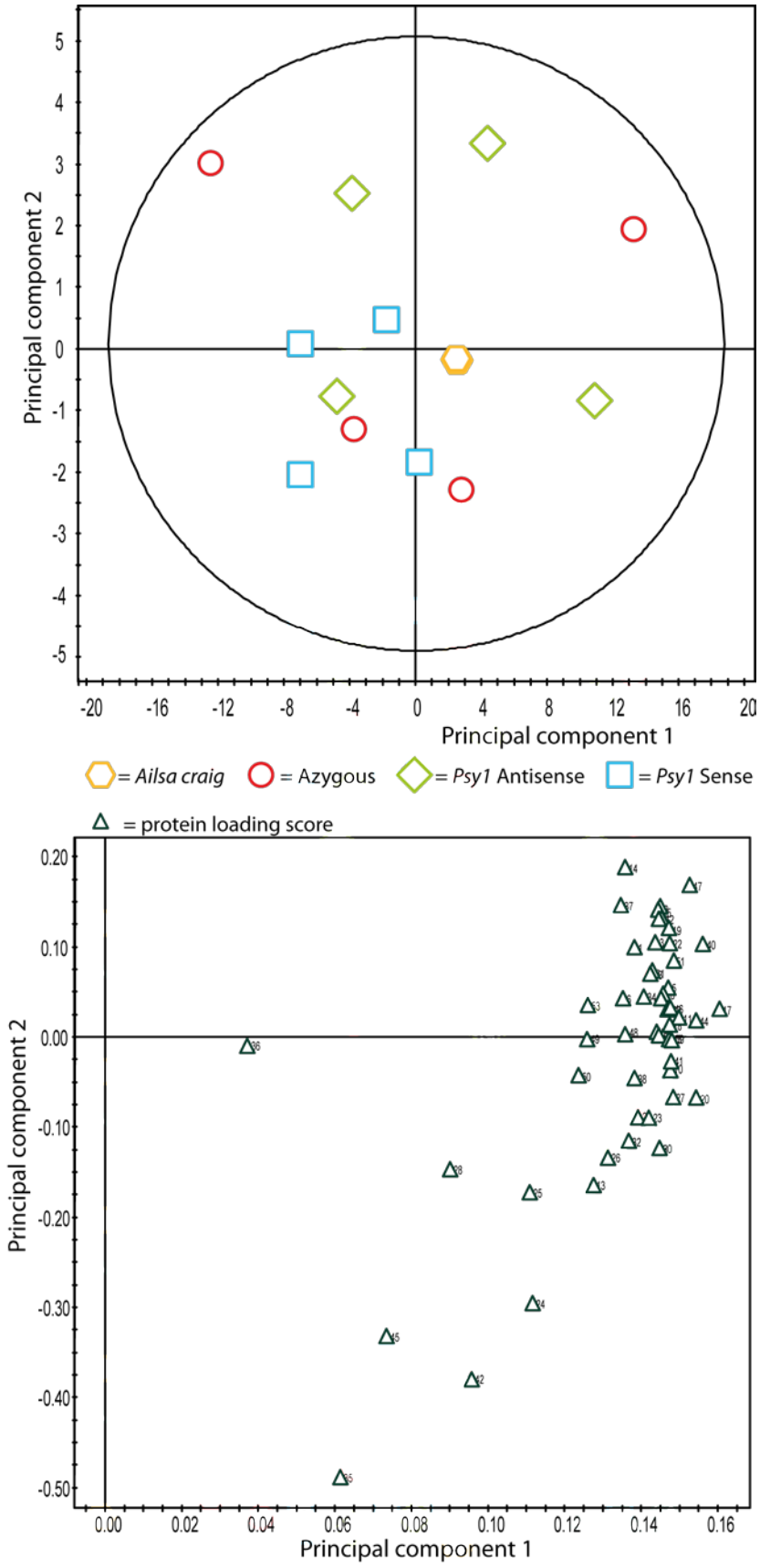


Figure 5.23 PCA of iTRAQ ratios from each cultivar

5.8.2.4 Variability between cultivars - Partial Least Square - Discriminant Analysis (PLS-DA)

Figure 5.24 displays the PLS-DA scatter plot and corresponding loadings plot for *Ailsa Craig* (AC), azygous (AZ), *Psy1* Antisense and *Psy1* Sense tomato cultivars. The cultivar coefficients do not form clusters and are interspersed across the plot. There is not enough variation between cultivars to allow classification of cultivar type from protein coefficients. The loadings score plot comprises protein coefficients. The protein coefficient for abscisic stress-ripening protein 1 is found close to the cultivar coefficient for *Psy1* Sense.

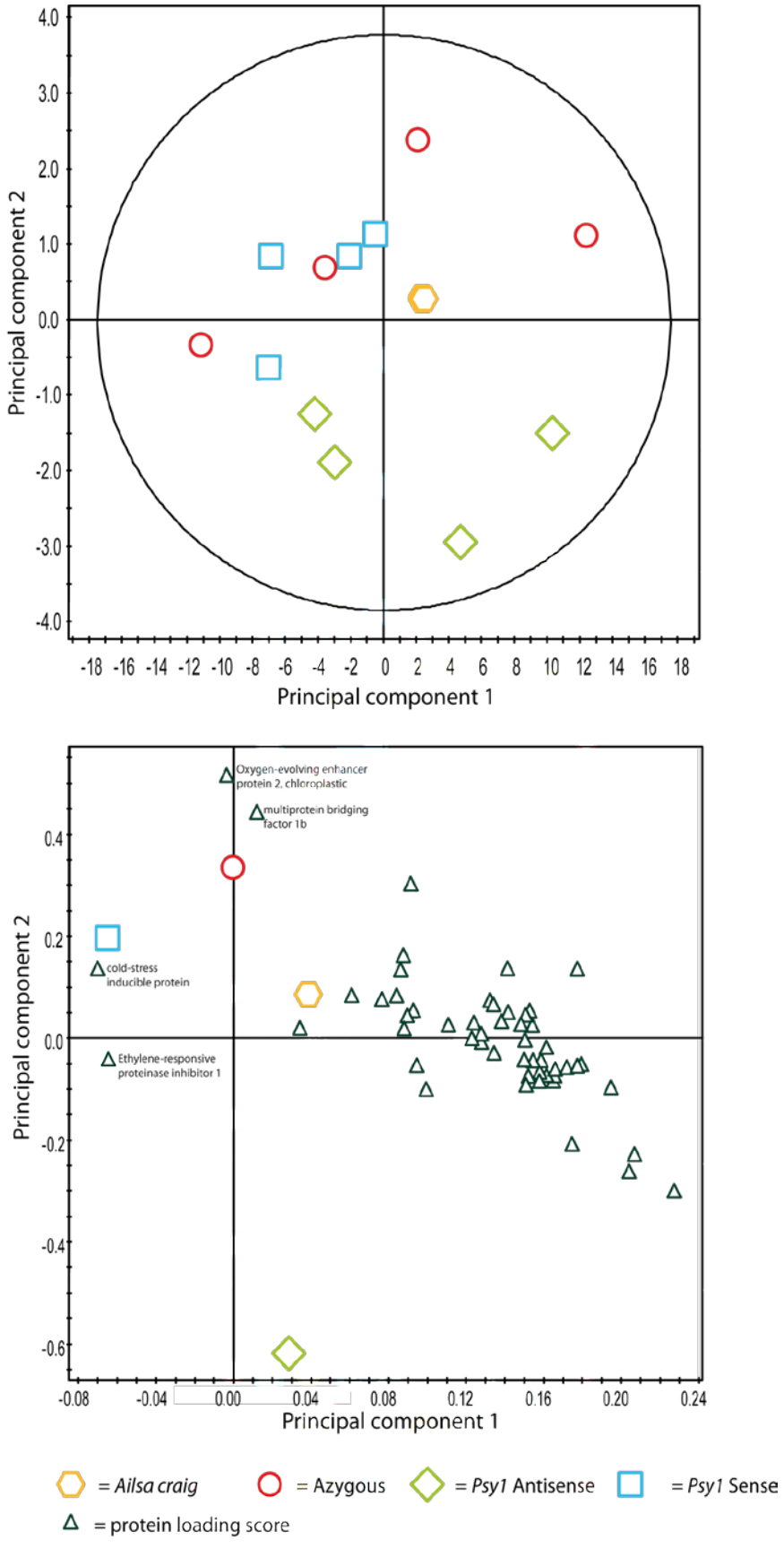


Figure 5.24 PLS-DA of iTRAQ ratios for each cultivar.

6 Discussion

6.1 MudPIT as a system for quantitative proteomics

During this project a MudPIT system was designed and developed for quantitative proteomics of GM crops. Its application to the assessment of GM crops has been demonstrated. In order to gain robust and reproducible data a variety of sample preparation methods were adopted to ensure optimal protein identification and coverage. Protein pre-fractionation techniques and sample preparation were found to be vital in developing a suitable workflow. For example the removal of lipids from samples and use of a urea based extraction buffer was essential to liberate membrane proteins and basic proteins.

Three different proteolytic enzymes were evaluated to find which gave the best protein coverage. Trypsin, Glu-C and Asp-N digestions were undertaken. Trypsin gave the highest number of proteins identified at 97 compared with 11 for Glu-C and13 in Asp-N tomato. Trypsin was used subsequently for all digestions. However, the use of multiple enzymes sequentially should not be ruled out as a strategy for enhancing coverage of proteomic and individual proteins.

A number of different pre-fractionation techniques were assessed. For example isoelectric focusing and BioRad OFFGEL systems were used. This allowed a large spectrum of proteins to be identified from tomato using subsequent digestion of proteins with trypsin and liquid chromatography coupled to a mass spectrometer. This method gave good coverage of proteins, especially from proteins with extremes of pH. However, these

systems had their drawbacks. Such as proteins being present in multiple fractions, major abundant proteins masking the presence of low abundant components and the processing time needed for a single sample was extended. Extended processing and analysis time was a limitation, as one sample was separated into many fractions, thus increasing analysis time and resources needed. As many proteins overlap several fractions it was difficult to concatenate data files after processing to get representative quantitative data.

Cation exchange chromatography, as a first dimension separation method, was tested on proteins extracted for MudPIT. Proteins were digested prior to separation then submitted for a 2^{nd} dimensional separation using reverse phase C_{18} chromatography coupled to a nano-LC and on-line mass spectrometer. The results revealed a good range of peptides and proteins. More proteins were identified than using the OFFGEL system, confirming MudPIT with cation exchange followed by C_{18} reverse phase nanoLC, could be used to compare proteins in a controlled experiment.

An optimal strategy and a workflow were adopted where proteins were extracted and digested with trypsin before separation with a cation exchange cartridge. One multiplexed fraction was separated into 4 cation exchange fractions that could be concatenated before data processing.

.

6.2 Assessing different MS platforms with the MudPIT system for quantitative proteomics

MudPIT experiments were undertaken on a variety of different MS platforms to compare their sensitivity and performance in protein

quantification. These were: Q-TOF, linear Ion Trap, Orbitrap and the MALDI-TOF-TOF platforms.

The Q-TOF, using the ABI QSTAR platform, identified 91 proteins in tomato and 42 from soya. The Q-TOF platform from Agilent identified 350 proteins from tomato. The QSTAR P*ulsar I* has a low duty cycle, compared to the Agilent 6520 machine which was able to scan faster to perform more MS/MS. The linear ion trap was not suitable for MudPIT experiments with iTRAQ, due to the one third cut off that limits detection of reporter ions essential for calculating peptide ratios. The MALDI-TOF/TOF was used with MudPIT soya proteins and identified 157 from non-GM soya. The results were comparable to results seen with the Q-TOF technology. The FT-Ion Trap Orbitrap gave excellent data and identified 374 proteins in tomato and 206 in soya.

Using the MudPIT approach a variety of protein products from gene manipulation were successfully identified in GM soya and GM tomato samples. EPSPS was successfully detected in GM soya MudPIT samples and bacterial phytoene desaturase was detected in GM tomato CRIT I, which is encouraging because they represent the intended effect of genetic engineering.

A study by Zybailov *et al*. (2009), developed a system for identifying 2904 proteins from leaf in *Arabidopsis* and assigning post-translational modifications. This study demonstrates the capabilities of third generation mass spectrometry platforms such as the LTQ Orbitrap XL and Velos machines and Bruker’s HCT ion traps and Q-TOF platforms.

6.3 MudPIT and gel based proteomics

Since the start of this project developments have been made in 2D gel proteomics. A 2D-DIGE approach was developed in 2006 by Stühler *et al.* for the Human Proteome Organisation Brain Proteome Project to monitor proteomic changes during murine brain development. This study found 420 differences, but highlighted the need to analyse more samples for statistical validation. 2D proteomic studies have been undertaken on GM plant systems. In 2008, a study by Zolla *et al.*, developed a proteomics 2D system to identify unintended side effects in transgenic maize as a result of GM. This study found that 100 total proteins were differentially modulated in the expression level as a consequence of the environmental influence and 43 proteins were found to be up- or down-regulated in transgenic seeds with respect to their controls. These changes were specifically related to the insertion of a single gene into a maize genome by particle bombardment. The transgenic seeds responded differentially to the same environment as compared to their respective isogenic controls, as a result of the genome rearrangement derived from gene insertion (Zollo *et al.*, 2008).

A 2D-PAGE proteomics study of tomato seedlings with salt stress was developed by Chen *et al.*, in 2009. This study was able to identify 400 features in the gels for comparison and quantification. This study identified 23 salt stress released proteins and used the technology to identify mechanism of salt stress resistance (Chen *et al.*, 2009).

In 2007 a system by Faurobert et al., has been developed to examine the proteome of tomato fruits during development using 2D GELS and MALDI-TOF MS for identification. This system identified 1,790 gel spot

features of which 90 proteins have been identified and quantified.

MudPIT experiments have been performed on plant and crop systems, for example the rice proteome. A comparison of gel-free MudPIT and 2-DGE revealed a much higher and unbiased representation of different protein classes by MudPIT (Rakwal & Agrawal, 2003).

Usaite et al. (2008) Multidimensional protein identification technology followed by quantitation using either spectral counting or stable isotope labeling approaches was used to identify relative changes in the protein expression levels between the strains. A total of 2388 proteins were relatively quantified. The stable isotope labeling based quantitative approach was found to be highly reproducible among biological replicates when complex protein mixtures containing small expression changes were analyzed. Where poor correlation between stable isotope labeling and spectral counting was found, the major reason behind the discrepancy was the lack of reproducible sampling for proteins with low spectral counts.

A study in 2010 by Barsan *et al.*, has identified 988 proteins from the tomato chromoplast. Proteins were sub-fractionated prior to analysis using sucrose density gradients. Proteins were extracted and analysed using an Orbitrap FT-Ion Trap MS/MS. This study shows that there is great promise and potential in gel based proteomics with label free quantification methods using fast scanning mass spectrometers.

6.4 Chemical vs. non-chemical quantification methods

This project aimed to compare chemical and non-chemical quantification systems with a variety of different MS platforms.

A non-chemical or label free method was developed with success using

the LCQ linear ion trap with MS^3 ions from iTRAQ. This method does not allow many proteins to be characterized in one experiment. The results, however, were highly accurate and the technique showed great promise for quantification of peptides.

A chemical quantification approach was designed and developed for soya and tomato systems using iTRAQ method. iTRAQ was deemed most suitable for GM crop assessment as the Q-TOF platform allowed quantification of 4 multiplexed samples of up to 300 proteins. There is further scope to use the FT-Ion trap Orbitrap platform to characterize over 400 proteins in a multiplexed sample.

Label free approaches have been applied to plant systems and nutritional sciences. In 2009 a study by Stevenson *et al.,* detailed seed allergens proteins that cause a majority of reported cases of food induced anaphylaxis. This study used peak integration and spectral counting with a linear ion trap mass spectrometer. Transgenic peanut lines, with a reduced allergen were tested against the wild type control and a reduction in allergenic proteins Ara h 2 (conglutin-7) and Ara h 6 (conglutin) were confirmed in the transgenic peanut line.

Bantscheff et al (2007) critically examine the more commonly used quantitative mass spectrometry methods for their individual merits and discuss challenges in arriving at meaningful interpretations of quantitative proteomic data and amopre recne treview is that of Zhu et al., 2010..

6.5 Establishing variation in tomato cultivars

iTRAQ experiments were undertaken on a wild type, azygous, and 2 GM cultivars to establish the variation between different plants of the same

cultivar. Using ANOVA, the number of proteins and peptides identified in each plant of cultivars were not significantly different from one another. The number of proteins identified when each sample was injected twice did not differ significantly. These results show that plants of the same cultivar give consistent variation in their protein complements and can be used to look at variation between plants from different cultivars.

Using the Agilent Q-TOF platform 350 proteins were identified in wild type cultivar in comparison to GM cultivar *Psy1* sense, where only 180 proteins were identified. The azygous variety did not significantly differ from the wild type indicating that genetic manipulation not the transformation process itself was responsible for changes. The *Psy1* Antisense cultivar did not show significant difference from the wild type. This cultivar silences the effect of the genetic manipulation in *Psy1* sense. As this tomato is not significantly different from the wild type, it would support the case that alterations in the gene product were responsible for differences seen in Psy1 sense cultivar.

A disappointing result from the iTRAQ experiments of different cultivars showed that the GM protein altered was not always identified in protein profiles. Thus the intended effect of genetic engineering was not always observed.

6.6 Examining variation between tomato cultivars

Results from the Students *t*-test give details of the variation to individual proteins in the tomato cultivars studied. In using the *t*-test it is possible to see which proteins are decreased or elevated to the wild type. In the azygous variety a total of seven proteins were found to be different. A

similar story was observed with *Psy1* antisense cultivar as ten proteins were perturbed, with changes existing in proteins involved in abiotic stress response.

In the *Psy1* sense cultivar over 60 of the 150 proteins examined were found to be perturbed. Tables 5.7 and 5.8 display proteins that were perturbed in *Psy1* sense cultivar, whilst Figures 5.21 and 5.22 show protein levels alongside metabolomic data. These indicate that many proteins found to have decreased levels in *Psy1* sense have metabolic functions, ie enzymes. The protein complement in *Psy1* sense is likely to be reduced as a result of phytohormone balance of the plant. The ripening profile of *Psy1* sense plants are altered by the genetic changes to levels of phytoene synthase. In *Psy1* sense phytoene synthase is over produced by the plant and GGPP is converted to phytoene from plant development. This intermediate precursor GGPP is utilised by other metabolic pathways including phytohormone like substances and in chlorophyll production. Phytoene synthase is over expressed in all plant tissues and GGPP is channelled to make phytoene at an early stage in plant development. The fruit have the same carotenoid profiles as wild type when they are red ripe. Many proteins are perturbed however, due to the up-regulation of one enzyme. In changing one enzyme other metabolites formed from the same substrate are altered at different stages of development and has a compromised system to cycle carbon (Fray *et al.*, 1995, Fraser *et al*, 2002)

6.7 Assessing substantial equivalence in genetically modified crops

The MudPIT experiments performed in this project have allowed

proteins to be relatively quantified from GM tomatoes and compared against the wild type control. The relative quantification method has allowed us to compare 150 proteins found in red ripe tomato fruits for the assessment of substantial equivalence. However, the gene product, phytoene synthase, altered by genetic manipulation was not identified for relative quantification in the MudPIT experiments performed.

By comparison of the azygous to wild type cultivar, it was possible to confirm that changes to protein complement were not due to plant transformation methods and regeneration. Comparisons of the protein complements of wild type and vs. *Psy1* Sense and *Psy1* Antisense cultivars indicated differences to the protein complements in both cultivars. In *Psy1* Sense cultivar 60 enzymes were down compared to the wild type. Figure 5.21 and 5.22 display protein levels in glycolysis, sucrose degradation and the TCA cycle pathways.

When these systems are examined at the pathway level, many changes and differences can be observed between wild type and GM cultivar. In the case of *Psy1* sense the changes indicate that many enzymes involved with basic carbon cycling processes such as glycolysis, sucrose degradation and TCA cycle are found at lower levels. However, these values do not estimate the enzyme activities of these proteins and therefore it is difficult to interpret protein changes as predictors of alterations to the metabolome. Similarly, changes to key metabolites may, in turn, cause further variaotions in the metabolome, especially enzymes, through post transcriptional events that have not been investigated in this project, In the case of *Psy1* sense, many pathways are affected by the changes to GGPP, which is used as a

metabolite in several reactions, including growth and signalling hormones. These plants have a disrupted ripening series and the changes could reflect senescence of plant tissues at an earlier stage. This implies that fruit need to be harvested at times post anthesis, rather than post breaker, as ripening per se can be affected by metabolite changes.

If applied to the substantial equivalence model, the *Psy1* sense cultivar shows many differences and changes to the wild type cultivar. Kuiper *et al.* (2001) recommended monitoring substantial equivalence to look for differences and similarities in GM cultivars. Once differences are identified they can be subject to further toxicological investigation.

In the case of *Psy1* sense, all enzymes examined were lower or equal to the levels in the wild type. One protein was slightly elevated, the abscisic stress ripening inhibitor. This protein is thought to be higher due to plant stress levels. Once again, however, one must be cautious in interpreting the significance of changes in protein levels without knowing enzymic activities and rates of reactions of rate-limiting steps. A multi-faceted approach to substantial equivalence needs to be taken in order to understand fully the alterations at protein and metabolite levels in GM crops.

7 Conclusions

- The MudPIT workflow is a suitable tool for examining proteins in tomato fruits and soya bean products.
- The MudPIT workflow can be applied to different Mass spectrometry platforms: ESI Q-TOF, Orbitrap and Linear Ion traps and MALDI-TOF/TOF.
- Relative quantification of 150 proteins can be achieved using the MudPIT workflow with iTRAQ chemical tags with a Q-TOF MS platform to assess protein perturbations in non-GM and GM tomato cultivars.
- In the azygous cultivar 7 proteins showed significant difference from the wild type, these proteins were stress responses.
- In the GM *Psy1* Sense cultivar, 60 proteins were found to be perturbed. 59 proteins were down regulated and one protein was found to be significantly elevated in *Psy1* Sense: abscisic acid stress ripening inhibitor protein 1.
- The gene product phytoene synthase from the intended genetic alteration in *Psy1* Sense was notably absent from the quantitative protein profile.

8 Acknowledgements

A great many thanks are owed to Douglas Lamont and Kenneth Beattie at the University of Dundee, Gary Woffendin in the demo labs of ThermoFisher Scientific in Hemel Hempstead and Rod Watson at Applied Biosystems in Warrington.

We thank the Food Standard Agency for financial support.

9 References

Baker, J.M., Hawkins, N.D., Ward, J.L., Lovegrove, A., Napier, J.A., Shewry, P.R. and Beale, M.H. (2006) A metabolomic study of substantial equivalence of field-grown genetically modified wheat. Plant Biotech. Journal 4 (4), 381–392.

Bantscheff . M Schirle, M Sweetman, G., Rick, J., Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. Anal. Bioanal. Chem. 389, 1017–1031.

Barsan, C., Sanchez-Bel1, P., Rombaldi, C., Egea, I., Rossigno, M., Kuntz, M., Zouine, M., Latche´, A., Bouzayen, M. and Pech, J-C. (2010) Characteristics of the tomato chromoplast revealed by proteomic analysis. J. Experimental Botany 61 (9), 2413–2431.

Boehm, A.M., Pütz, S., Altenhöfer, D., Sickmann, A. and Falk, M. (2007) Precise protein quantification based on peptide quantification using iTRAQ. BMC Bioinformatics 8 (214), 1-18.

Catchpole, G.S., Beckmann, M., Enot, D.P., Mondhe, M., Zywicki, B., Taylor, J., Hardy, N., Smith, A., King, R.D., Kell, D.B., Fiehn, O., Draper, J. (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. PNAS 102 (40), 14458-14462.

Cellini F., Chesson A., Colquhoun I., Constable A., Davies H.V., Engel K.H., Gatehouse A.M.R., Kaerenlampi E.J., Leguay J.-J., Legesranta S., Noteborn H.P.J.M., Pedersen J., Smith M. (2004). Unintended effects and their detection in genetically modified crops. Food and Chemical Toxicology 42, 1089-1125.

Chen, E.I., Hewel, J., Felding-Habermann, B. and Yates, J.R. (2006) Large scale protein profiling by combination of protein fractionation and

multidimensional protein identification technology (MudPIT). Mol. Cell. Proteomics 5, 53-56.

Chen, S., Gollop, N. and Heuer, B. (2009) Proteomic analysis of salt-stressed tomato (Solanum lycopersicum) seedlings: effect of genotype and exogenous application of glycinebetaine. J. of Exp. Botany 60 (7), 2005-2019.

Coll, A., Nadal, A., Collado, R., Capellades, G., Kubista, M., Messeguer, J., and Pla, M-J. (2010) Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices. Plant Mol. Biol.73 (3), 349-362.

Faurobert, M., Mihr, C., Bertin, N., Pawlowski, T., Negroni, L., Sommerer, N. and Causse, M. (2007) Major proteome variations associated with cherry tomato pericarp development and ripening. Plant Physiology 143, 1327–1346.

Fenyö, D. and Beavis, R.C. (2008) Bioinformatics development: Challenges and solutions for mass spectrometry. Mass Spectrom. Revs. 27, 1-19.

Fernández-Ocaña, M., Fraser, P.D., Patel, R.K., Halket, J.M., Bramley, P.M. (2007) Mass spectrometric detection of CP4 EPSPS in genetically modified soya and maize. Rapid Commun Mass Spectrom. 21 (3), 319-328

Fraser, P.D., Romer, S. Shipton, C.A., Mills, P.B., Kiano, K.W., Misawa, N., Drake, R.G., Schuch, W., Bramley, P.M. (2001) Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner. PNAS 9 (2), 1092- 1097.

Fray, R.G., Wallace, A., Fraser, P.D., Valero, D., Hedden, P., Bramley, P., Grierson, D. (1995) Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the giberillin pathway. The Plant J. 8 (5), 693-701.

Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. and Gypi, S.P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem mass spectrometry. PNAS 100 (12), 6940-6945.

Gonzalez-Begne, M., Lu, B., Han, X., Hagen, F.K., Hand, R.A., Melvin, J.E., Yates, J.R. (2009) Proteomic analysis of human parotid gland exosomes by multidimensional protein identification technology (MudPIT). J. Proteome Res. 8 (3), 1304–1314.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotech. 17, 994-999.

Halket, J.M., Waterman, D., Przyborowska, A.M., Patel, R.J.K., Fraser, P.D., Bramley, P.M. (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J. Exp. Bot. 56, 219-243.

Herrick, C.B. (2005) ‘Cultures of GMO’:discourses of risk and labelling of GMOs in the Uk and EU. Area. 37 (3), 286-294.

Hossain, F. and Onyango, B. (2004) Product attributes and consumer acceptance of nutritionally enhanced genetically modified foods Int. J. of Consumer Studies. 28 (3), 255–267.

Koller, A., Washburn, M.P., Lange, B.M., Andon, N.L., Deciu, C., Haynes, P.A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters, D. and Yates III, Y.R. (2002) Proteomic survey of metabolic pathways in rice. PNAS. 99 (18), 11969-11974.

König, A., Cockburn, A., Crevel, R.W.R., Debruyne, E., Grafstroem, R., Hammerling, U., Kimber, I., Knudsen, I., Kuiper, H.A., Peijnenburg, A.A.C.M., Penninks, A.H., Poulsen, M., Schauzu, M., and Wal, J. M. (2004) Assessment of the safety of foods derived from genetically modified (GM) crops. Food and Chem. Tox. 42 (7), 1047-1088.

Kuiper, H.A., Kleter, G.A., Noteborn, H.P.J.M., Kok, E.J. (2001) Assessment of the food safety issues related to genetically modified foods. The Plant Journal. 27(6), 503–528.

Moffitt Cancer Center and Research 2009 (2009) http://www.moffitt.org/Site.aspx?spid=FE4E515F82C14C0BB40648CA1E27465F,15th February 2010.

Rakwal, R. and Agrawal, G.K. (2003) Rice proteomics: Current status and future perspectives. Electrophoresis 24, 3378–3389

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D.J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Prot. 3 (12), 1154-1169.

Stühler, K., Pfeiffer, K., Joppich, C., Stephan, C., Jung, K., Müller, M., Schmidt, O., van Hall, A., Hamacher, M., Urfer, W., Meyer, H.E., Marcus, K. (2006) Pilot study of the Human Proteome Organisation Brain Proteome Project: Applying different 2-DE techniques to monitor proteomic changes during murine brain development. Proteomics 6 , 4899-4913.

Usaite R., Wohlschlegel J., Venable, J.D., Park, S.K., Nielsen, J., Olsson, L., Yates III J.R. (2008) Characterization of Global Yeast Quantitative Proteome Data Generated from the Wild-Type and Glucose Repression *Saccharomyces cerevisiae* Strains: The

Comparison of Two Quantitative Methods. *J. Proteome Res.*, **2008**, ,*7* 266–275

Washburn, M.P., Wolters, D., Yates, J.R.III. (2001) Large-scale analysis of the yeast proteome via multidimensional protein identification technology. Nat Biotech. 19, 242-247.

Whitelegge, J.P. (2005) Mass spectrometry for high throughput quantitative proteomics in plant research: lessons from thylakoid membranes Plant Physiol. Biochem. 42, 919-927.

Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D.A. and White, F.M. (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. PNAS. 104 (14), 5860-5865.

Wolters, D.A., Washburn, M.P., and Yates, J.R.III and Wolters, D. (2001) An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem 73, 5683-5690.

Zhu, W., Smith,J.W., Huang, C-M. (2010) Mass Spectrometry-Based Label-Free Quantitative Proteomics. Journal of Biomedicine and Biotechnology , doi:10.1155/2010/840518

Zolla, L., Rinalducci, S., Antonioli, P. and Righetti, P.G. (2008) Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a Result of genetic modifications. J. Proteome Res., 7 (5), 1850-1861.

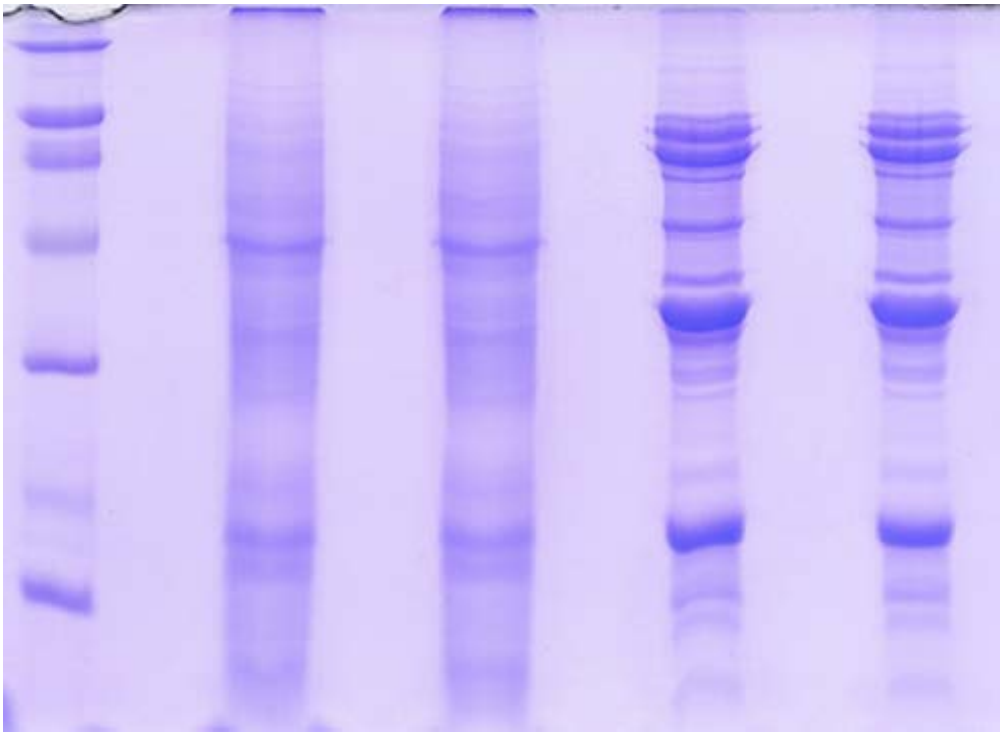
Zybailov, B., Sun, Q. and van Wijk, K.J. (2009) Workflow for large scale detection and validation of peptide modifications by RPLC-LTQ-Orbitrap: Application to the *Arabidopsis thaliana* leaf proteome and an online modified peptide library. Anal. Chem. 81 (19), 8015-8024.

Appendix

Objective-02, tasks 02/01-02. To establish a reproducible, efficient extraction procedure for proteins

Plant material (ripe tomato fruit) were homogenised with freezer mill, which gave the most effective homogenisation and different solutions and extraction conditions were optimised to maximise protein solubilisation. The best solubilisation was achieved using a solution containing 7M Urea, 2M thiourea, 2% CHAPS and 10 mM DTT when extraction was performed for 30 min in shaking in RT (Figure 1). Proteins were precipitated using 5 × vol of ice cold acetone in -20°C overnight.

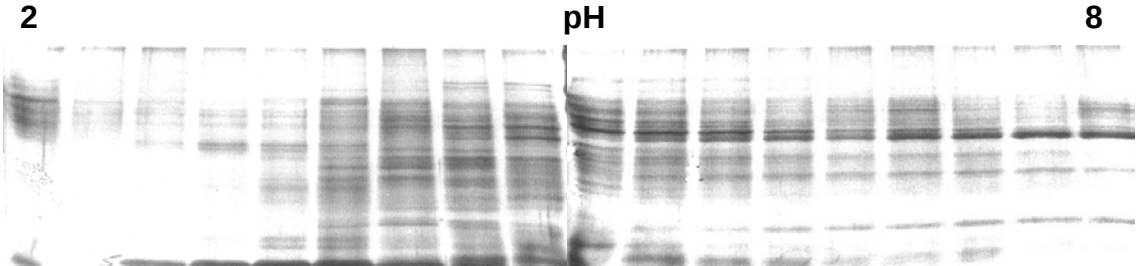
Figure 1 Coomassie stained gel of fractions from



Objective 3, task 03/01-03. Pre-fractionation/purification methods (subcellular fractionation, liquid isoelectric focusing and anion-exchange chromatography) were evaluated.

1.1. Membrane and soluble fractions fo toamot fruit were separated using ultracentrifugation (105,000g)in the abscoeb iof urea. Liquid isoelectring focusing (Rotofor) gave good enrichment of proteins and decreased the complexicity of sample, however, the most abundant proteins were found from several fractions (Figure 2). Furthermore, the method is not suitable for high-throughput analysis.

Figure 2. Profile of protein bands following separation in a Rotofor.



Objective 4, task 04/01-02.

Trypsin was found to be the most effective digestion enzyme compared to Glu-C and Asp-N. Figure 3 presents a table containing total amount of identified proteins and protein coverage of some selected proteins. However, the results indicate that different enzymes give complementary results and the other enzymes could be used for specific applications.

Efficiency of different proteolytic enzymes and sequence coverage of some selected proteins. Number of identified proteins is obtained by Mascot.

Protein	Accession number	Sequence coverage (trypsin)	Sequence coverage (Glu-C)	Sequence coverage (Asp-N)
Tomato		**97 proteins**	**11 proteins**	**13 proteins**
Abscisic stress ripening protein 1	gi|584786	(18/115 = 15.7%)	(33/115 = 28.7%)	(0/115 = 0%)
coat protein	gi|229181	(120/158 = 75.9%)	(42/158 = 26.6%)	(32/158 = 20.3%)
Acid beta-fructofuranosidase precursor (Acid sucrose hydrolase)	gi|124701	(131/636 = 20.6%)	(29/636 = 4.6%)	(163/636 = 25.6%)
1-aminocyclopropane-1-carboxylate oxidase homolog (Protein E8)	gi|119640	(94/363 = 25.9%)	(0/363 = 0%)	(115/363 = 31.7%)
Soya		**32 proteins**	**14 proteins**	**13 proteins**
glycinin G4 subunit	gi|255224	(187/560 = 33.4%)	(0/560 = 0%)	(0/560 = 0%)
alpha subunit of beta conglycinin [Glycine max]	gi|9967361	(126/559 = 22.5%)	(170/559 = 30.4%)	(99/559 = 17.7%)
Sucrose-binding protein precursor (SBP)	gi|548900	(54/524 = 10.3%)	(66/524 = 12.6%)	(0/524 = 0%)
24 kDa oleosin isoform [Glycine max]	gi|266689	(13/223 = 5.8%)	(28/223 = 12.6%)	(10/223 = 0%)
P34 probable thiol protease precursor	gi|129353	(0/379 = 0%)	(33/379 = 8.7%)	(19/379 = 5.0%)

Objective 5, task 05/01-04. Establish reproducible off- and on-line multidimensional separation procedure for MudPIT

- Off-line cation-exchange method was further optimised for both tomato and soya, and good separation of peptides was obtained
- For the analysis of strong cation exchange (SCX) fractions, the RP nano-LC method was developed for ESI-qTOF-MS/MS instrument. Good separation of peaks was achieved. (Figure 2).

- Nano-LC system with monolithic nano columns was developed. With monolithic trap column and analysis column the resolution of peaks was better than when using conventional C18 columns (Figure 3).
- Capillary LC method with monolithic capillary columns was used for ESI-MS^n (ion-trap) instrument (Figures 4) to analyse SCX fractions of tomato and soya. Good separation and intensity was achieved.

Fig 2 Cation-exchange chromatograms of trypsin digested tomato and soya proteins. Profile of gradient is indicated as red lines. Flow was 0.2 ml/min and fraction were collected every 2 min.
Column: PolySULFOETHYL A, 200 x 2.1 mm, 5 µm, 200 Å (The Nest Group)
Solvents: A = 10 mM KH2PO4, 25% ACN, B = 10 mM KH2PO4, 25% ACN, 1 M KCl

Gradient: 0-20 min 0% B; 20-57.5 min 0-15% B; 57.5-70 min 15-50% B; 70-80 min 50% B; 80-90 min 50-0% B; 90-110 min 0% B

Tomato Soya

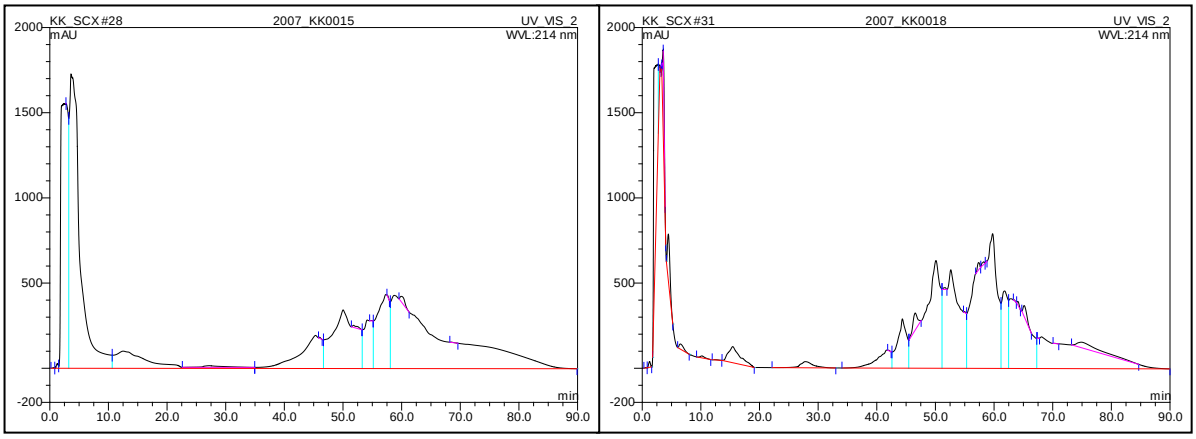
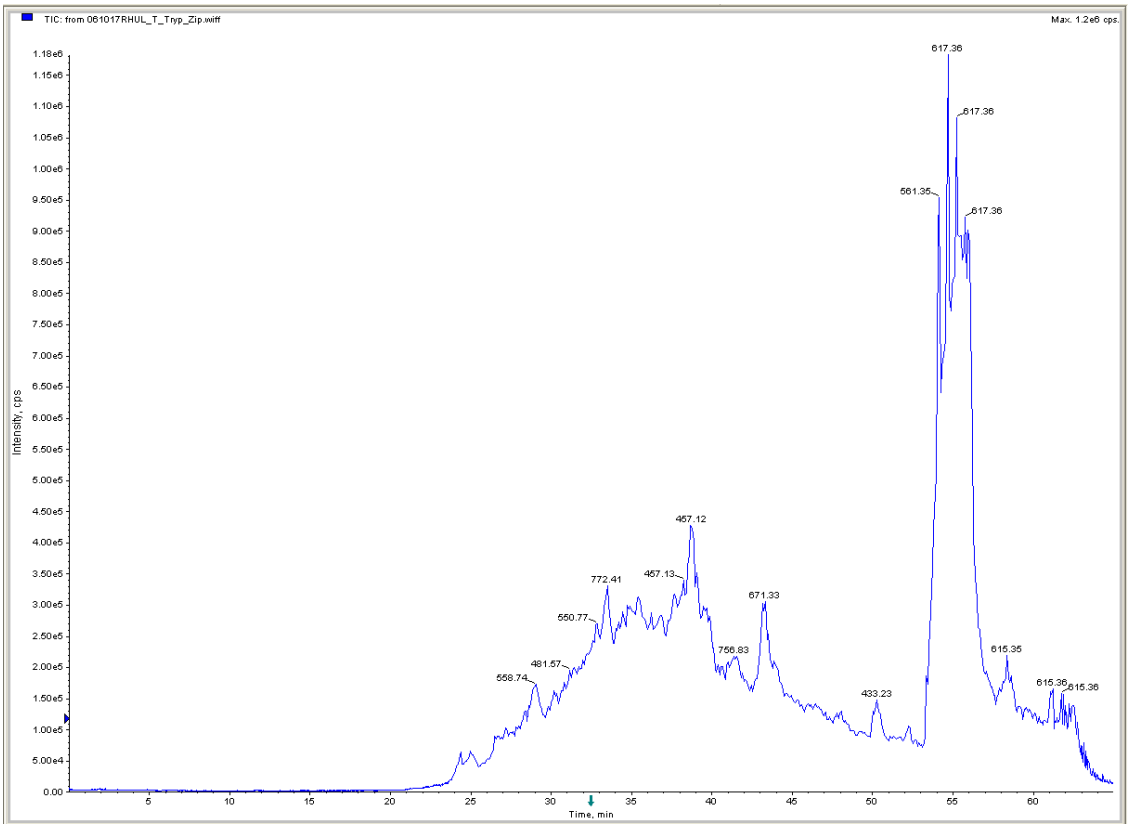


Fig. 3RP-nanoLC TIC spectrum of trypsin digested tomato sample.
Column: PepMap C18, 75 µm, 15 cm, 3 µm, 100 Å (Dionex)
Solvents: A = 0.1% FA, 5% ACN, B = 0.1% FA, 95% ACN, Loading solvent = 0.1% FA, 2% ACN Gradient: 0-40% B in 40 min, flow 200 nl/min

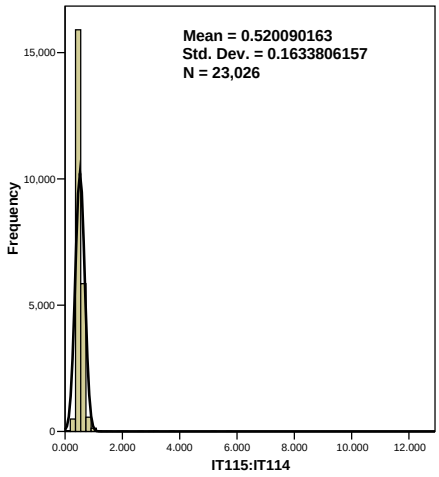
2. 
TIC: from 061017RHUL_T_Tryp_Zip.wiff
Max. 1.2e6 cps.
Intensity, cps
Time, min
617.36
617.36
561.35
617.36
457.12
457.13
772.41
671.33
550.77
481.57
558.74
756.83
433.23
615.35
615.36
615.36

Objective 7, task 07/01.

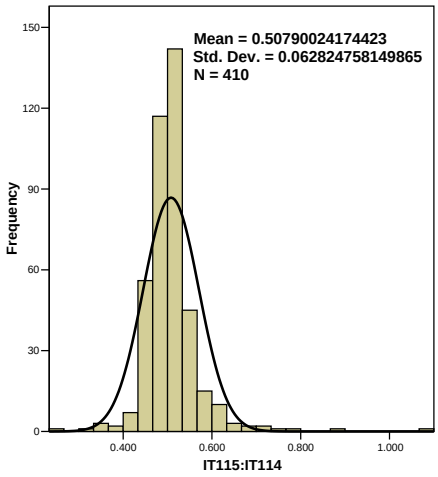
iTRAQ as multiplex quantitative MudPIT approach was tested in Applied Biosystems facility in Warrington. 100% and 50% samples of both tomato (CrtI) and soya (control) were tested.

- Developed and optimised extraction method and digestion conditions were compatible with quantitative MudPIT and 410 unique proteins (~23000 peptides) were identified.
- In first experiment 95% of peptides in soya and 85% of peptides in tomato were labelled. Analysis of tomato samples was repeated using double amount of iTRAQ reagent and 100% labelling efficiency was obtained.
- Quantitation within run was good and 90% of peptides gave 50%/100% between 0.395 and 0.681 and 90% proteins the values were between 0.435 and 0.603

Peptides:



Proteins:



Histograms of quantified peptides (n~ 23000) and proteins (n = 410). For the peptides 90% of values are between 0.395 and 0.681 and for proteins 90% of values are between 0.435 and 0.603.

Overviews of the iTRAQ procedure